Disproportionality

Data Analytics Lab

January 2020

This document summarises the assessment of disproportionality across several core WMP systems. Disproportionality is measured by estimating the relative risk of groups of nominals on an outcome, given a common characteristic of the nominals. The focus of this report will be on the nominal ethnicity, age and sex.

For each system, data has been selected to include key characteristics such as ethnicity and sex as well as other potentially crucial characteristics such as outcome of Stop and Search or role type in crime. Tables have been constructed counting the number of (distinct) nominals in each outcome across two characteristics (for example, male and female) see this section for details of the relative risk methodology. This is usually done via the comparison of the data in one system to the 'population' which nominals in that system originate from; finding an appropriate 'population' is not always straight forward and for this reason three population types are used.

1 Table of Contents

2 Introduction

The brief is to examine WMP data for evidence of disproportionality, in particular by ethnicity and sex. There is already a lot of work which aims to measure disproportionality in some manner, however, this work intends to include several systems which are not always included such as custody and crimes. This analysis will also include the comparison of multiple base populations to provide a more rounded view of disproportionality.

In accordance with general practice relative risk is used here rather than an absolute risk because the aim of this analysis is to give an indication of the effect of certain characteristics on an outcome between different groups of people. The 95% confidence interval is also given for each relative risk to indicate the level of uncertainty of an estimate.

Relative risk is defined here as (using Stop and Search as an example):

$$\frac{\left(\frac{No.\ white\ stopped\ \&\ searched}{White\ population}\right)}{\left(\frac{No.\ non-white\ people\ stopped\ \&\ searched}{Non-white\ population}\right)}$$

Whereby the relevant population could be different from the population as a whole. See the appendix for details.

3 Analysis of the Population

The 2011 Census data has been obtained to understand the population, ethnicity and sex splits across the WMP region. The counts of nominals residing within the WMP region by ethnicity and sex, can then be compared to the counts of nominals in each system.

3.1 Data

Several datasets are used in the analysis of the population, most of which are spatial units:

- The first contains population, NPU, crime rate and geographical information for all (WMP defined) neighbourhoods in the WMP region.
- The second contains population and geographical information for all WMP Neighbourhood Policing Units (NPUs).
- The 2011 Census data contains counts per ethnicity, sex, age group, total population and employment as well as other metrics for each of the lower layer super output areas (LSOA) (National Statistics 2012) in the WMP region[1].
- The fourth contains the geometries for each LSOA in England and Wales.
- The last contains the Index of Multiple Deprivation (IMD) data from 2010[2] for each LSOA in England and Wales, this includes the IMD score and rank, lowest being the most deprived LSOA (National Statistics 2011).

3.2 Outputs

The number of nominals residing in each LSOA within the WMP region is then aggregated for each characteristic to the whole WMP region. This then produces the number of nominals in each ethnic group, each sex and each NPU in the WMP region. These tables are then saved for later comparisons. The three tables are given below.

Ethnicity	Number of Nominals	Percentage
white	1,919,138	70.13%
mixed	96,204	3.52%
asian	514,981	18.82%
black	164,069	6%
other	42,068	1.54%

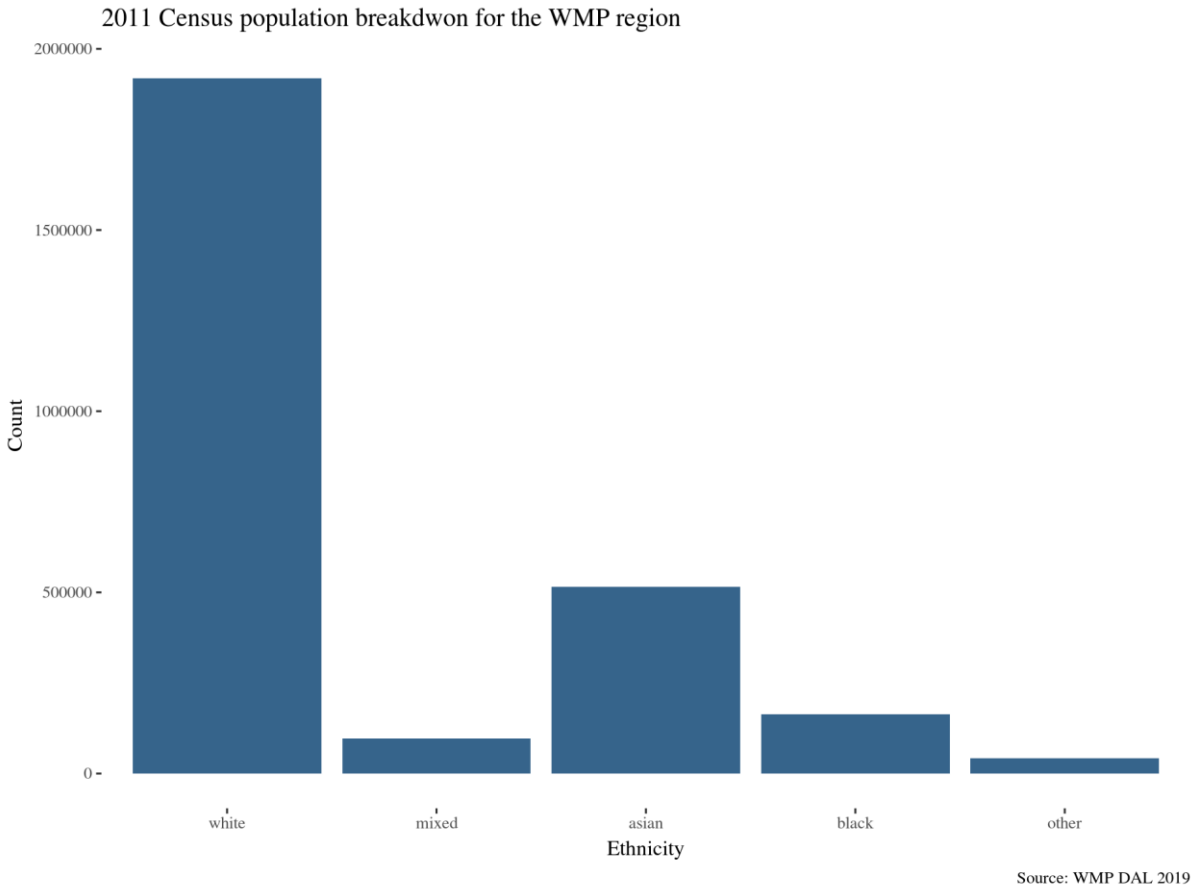
[1] Output areas are spatial units that have been created by ONS for the purposes of statistical reporting (their boundaries are not supposed to change allowing comparison through time), although see the note below.

[2] The 2011 Census data is the latest in the series of Census data, there is a more recent IMD, however, after 2011 the LSOA boundaries changed and so the 2015 IMD LSOAs do not match the 2011 Census. Hence here the 2011 Census and 2010 IMD are used.

Ethnicity	Number of Nominals	Percentage

Number of nominals of each ethnic group residing in the WMP region

A plot of this data is also provided for comparison with ethnicity breakdowns from other systems.



Sex	Number of Nominals	Percentage
male	1,346,950	49.22%
female	1,389,510	50.78%

Number of nominals of each sex residing in the WMP region

These splits differ across the NPUs.

NPU	NPU Code	Number of Nominals	Percentage
Birmingham East	BE	536,564	19.61%
Birmingham West	BW	536,474	19.6%
Coventry	CV	316,959	11.58%
Dudley	DY	312,929	11.44%
Sandwell	SW	308,064	11.26%

NPU	NPU Code	Number of Nominals	Percentage
Solihull	SH	206,678	7.55%
Walsall	WS	269,339	9.84%
Wolverhampton	WV	249,456	9.12%

Number of nominals residing in each NPU in the WMP region

Work undertaken elsewhere within WMP has delineated areas of (relatively) high demand for WMP's services which generally coincide with higher ranking IMD areas. To take this point into consideration, the 2011 Census data is joined to the IMD data which is then filtered to ten percent of the most deprived LSOAs, the mean proportions for each ethnic group are taken and then scaled to the population of the WMP region as a whole. This has been done to try and take locational characteristics and the geographic focus of much of WMP's demand into account that may affect the output. The outputs of this approach are given below.

Ethnicity	Nominals in Deprived Areas	Population Proportion	Deprived Proportion	Percentage Change
white	132,721	0.700	0.460	-34%
mixed	15,643	0.040	0.050	42%
asian	94,548	0.190	0.330	75%
black	39,455	0.060	0.140	134%
other	8,119	0.020	0.030	95%

Ethnic group proportions from the top ten percent of the most deprived areas scaled to the population of the WMP region

Sex	Nominals in Deprived Areas	Population Proportion	Deprived Proportion	Percentage Change
male	144,863	0.490	0.500	1.58%
female	145,623	0.510	0.500	-1.53%

Sex proportions from the top ten percent of the most deprived areas scaled to the population of the WMP region

4 Crimes

The Crimes system contains all details relating to crimes and non-crimes and the nominals involved, this could include victims, offenders, suspects and other roles such as siblings, person offence reported to and parent or guardian. The Crimes database has been queried to obtain nominal and crime details for analysis.

4.1 Data

After a series of joins, the fields returned from the Crimes database are:

Field	Data Type	Description
crime_ref	int	Crime reference number
offence_number	character	Offence reference
date_record_created	datetime	Date record created in Crimes system
offence_title	character	Description of offence
description	character	Detailed role description
role_type	character	Role type
nominal_ref	character	Nominal reference
sex	character	Nominal sex
date_of_birth	date	Nominal date of birth
ethnicity_detail	character	Ethnicity description
age_at_offence	int	Nominal age at date_record_created
ethnicity	character	Nominal ethnic group
address_ref	int	Address reference
curr_lpu	character	NPU where offence was committed
city	character	City where offence was committed
status	character	Whether the nominal has a victim or offender role
crime_class	character	Catagory of offence

Data dictionary of Crimes data returned

For the rest of this analysis, a nominal is included if their role in a record was one of the following:

- “SUSPECT”
- “DEFENDANT/OFFENDER”
- “PERSON REPORTED FOR CRIME/OFFENCE”

- "PERSON THOUGHT RESPONSIBLE FOR THE OFFENCE"
- "PERSON PROBABLY RESPONSIBLE"
- "POTENTIAL DETECTION".

A nominal is considered to be a victim if their role in a record was "VICTIM" or "ADDITIONAL VICTIM".

Due to requiring consistency with Census data, some ethnicities were grouped prior to the analysis, the mapping used with the number of nominals per ethnciity before aggregating is defined below.

Ethnicity before mapping	Ethnicity after mapping	Count nominals before mapping (in Crimes)
WHITE NORTH EUROPEAN	WHITE	1,465,689
ASIAN	ASIAN	393,841
BLACK	BLACK	250,203
NOT KNOWN	NOT KNOWN	184,145
OTHER	OTHER	49,669
WHITE SOUTH EUROPEAN	WHITE	19,650
MIDDLE EASTERN	OTHER	9,103
CHINESE/JAPANESE/SOUTH EAST ASIAN	CJSEA	8,603
CHINESE	CJSEA	869
BANGLADESHI	ASIAN	239

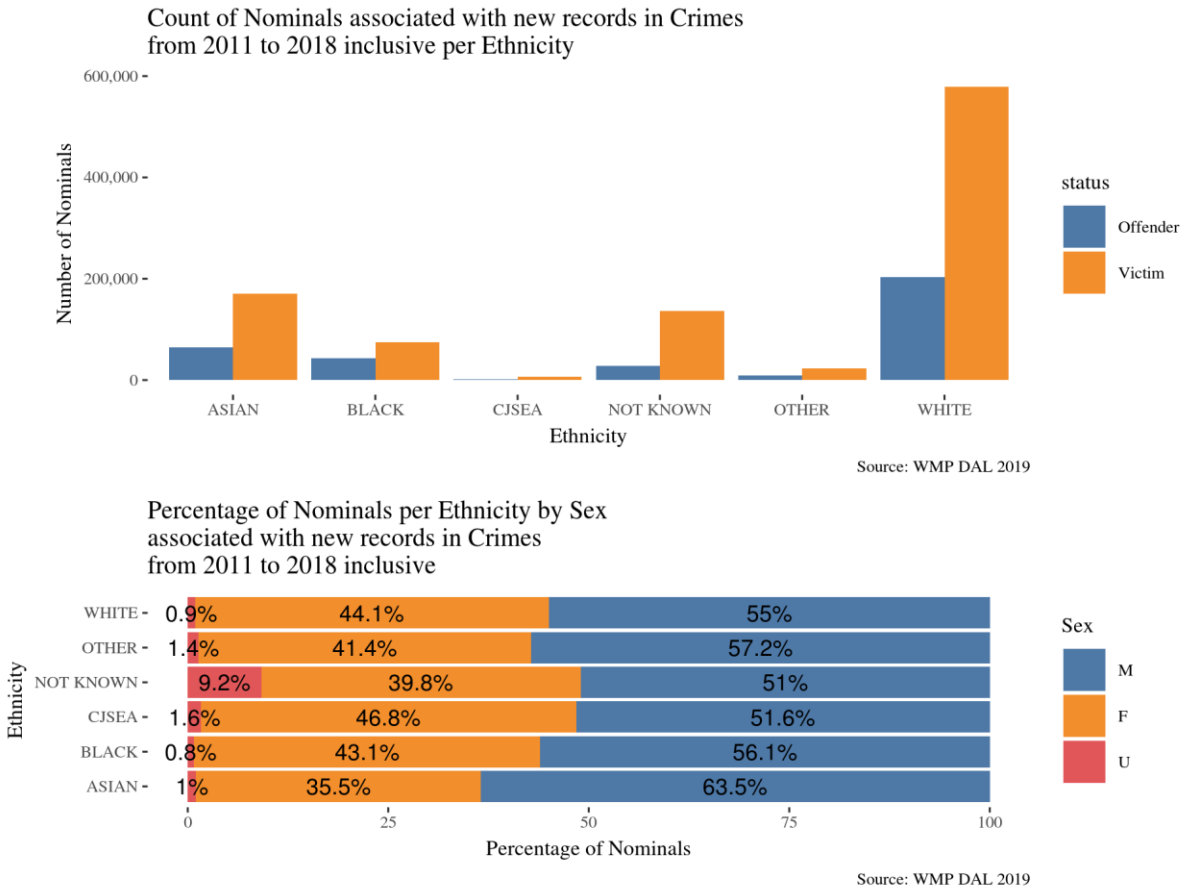
4.2 Results

For the majority of the outputs, a contingency table is constructed from the Crimes data for an outcome (e.g. offender) across a characteristic (e.g. ethnicity). This is then compared to several estimates of the population across the same characteristic to calculate estimates of relative risk. The difficulty here is the "population"; there is no one true answer for the population from which offenders or victims originate. One estimate could be the population of the WMP region; however, there is also the population of locations from which the majority of demand for WMP's services arises.. Also, the offenders in Crimes only include those that have been identified. There are most likely crimes with no offender identified but in reality there is one, and they could be a part of the population. However, there is no simple solution for this. The latest Census data is from 2011, hence here an estimate of the 2018[3] population was also used.

[3] Population estimates for years following a Census are produced by the ONS. These include total population and population by age, but not for sex or ethnicity. These have been derived by way of a two point exponential population growth model applied to the years 2001 and 2011 for each ethnicity / sex

Below are a series of plots summarising the analysis:

The first plot shows the count of the number of distinct nominals, distinctness is determined by the `nominal_ref`, per ethnicity that are associated with new records in the Crimes system between 2011 and 2018 inclusive. This includes both offenders and victims. The second half of the plot shows the percentage of distinct nominals of each sex per ethnicity.



Two observations from the above plots:

- The ethnicity breakdown is similar to that of the population ethnicity breakdown.
- There is a higher percentage of male nominals in Crimes, particularly in the Asian ethnic group, compared to the wider population.

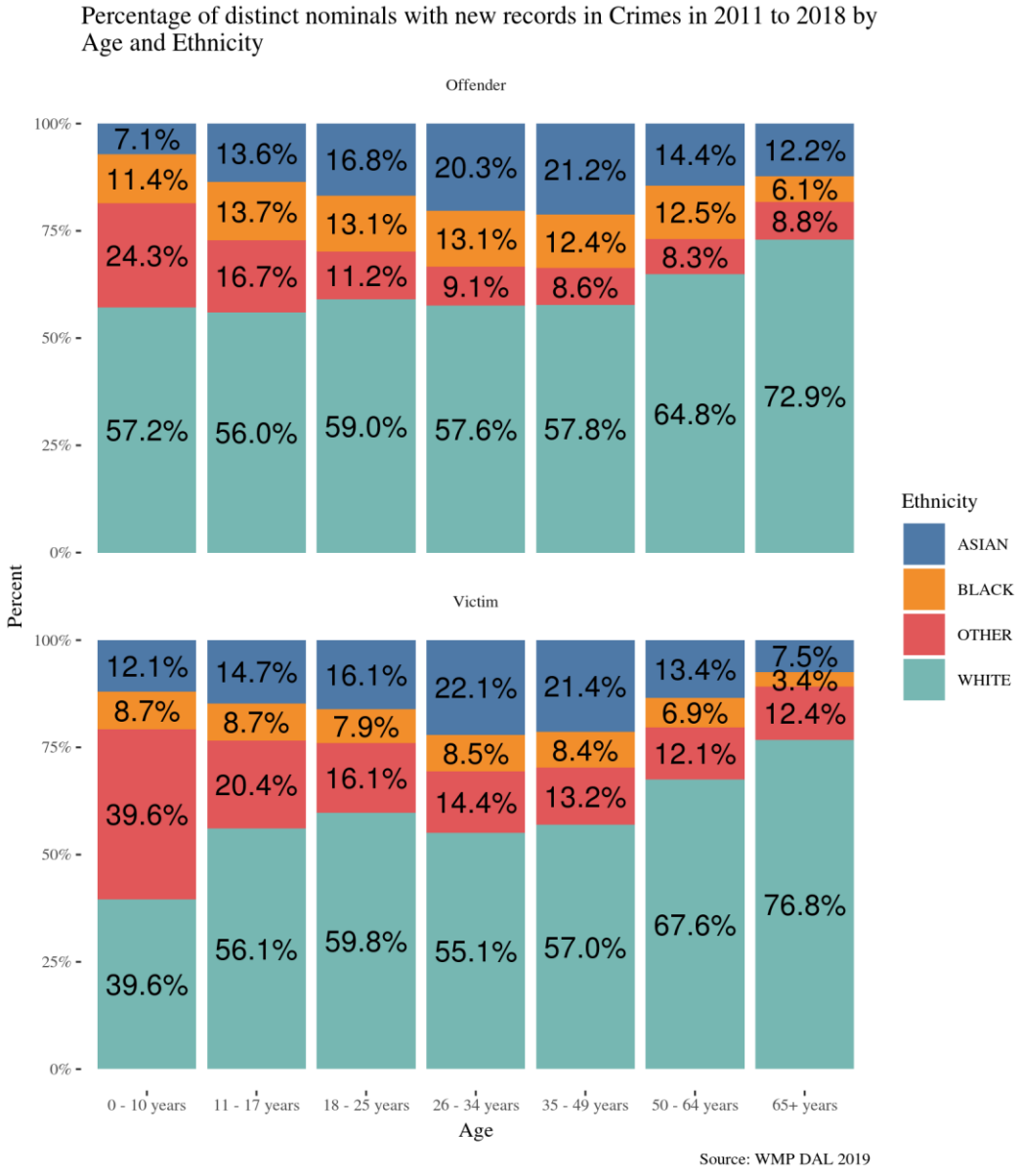
The plot below shows the percentage of distinct nominals of Asian, Black, White and other (CJSEA, other and not known combined) ethnic groups per age group.

- In both halves of the plot the percentage of distinct white nominals increases as the age group increases.
- In both halves of the plot the percentage of distinct black nominals remains almost constant, although with a drop in the older age band.

---

with the resulting proportionate presence of the different ethnicities then applied to the ONS estimate of total 2018 population.

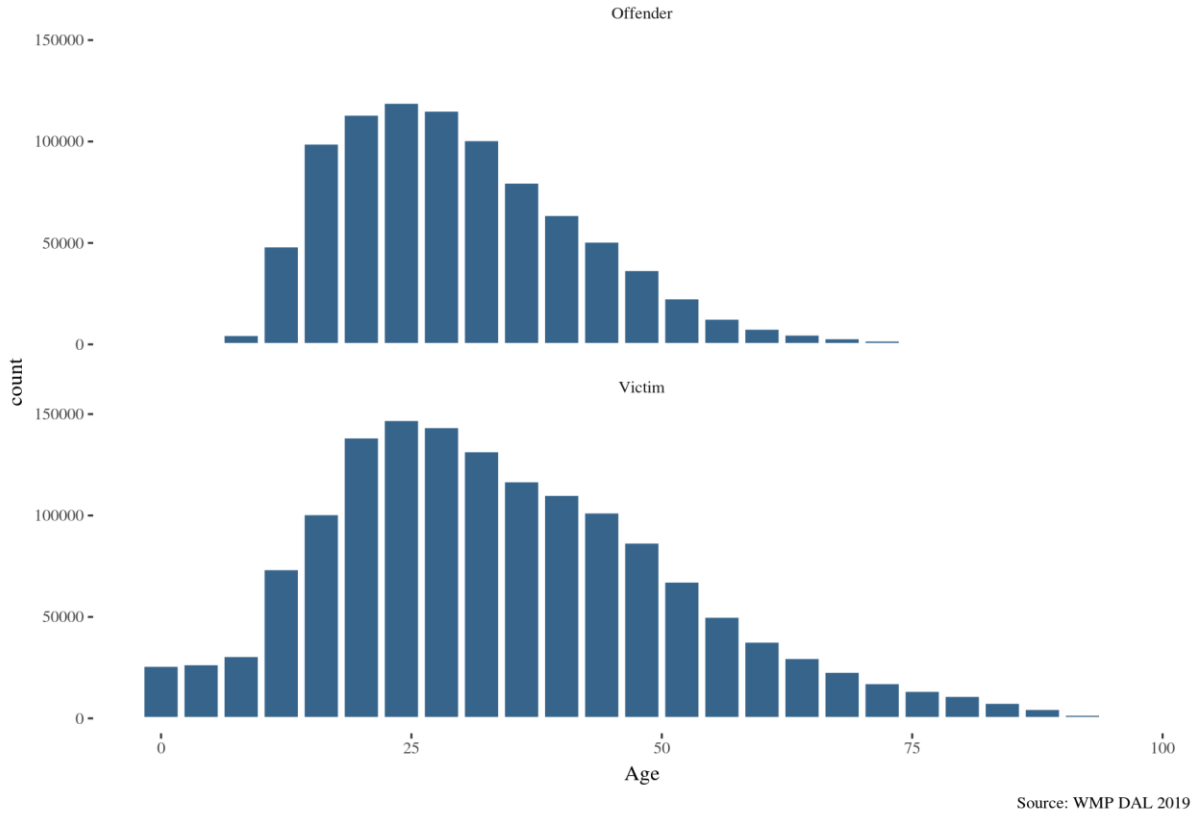
- In both halves of the plot the percentage of distinct asian nominals peaks between 26 and 49 years old and is relatively constant for all other age groups.
- In both halves of the plot the percentage of distinct other nominals decreases as the age group increases.



The next plot shows a histogram of the nominals' age for both offenders and victims.

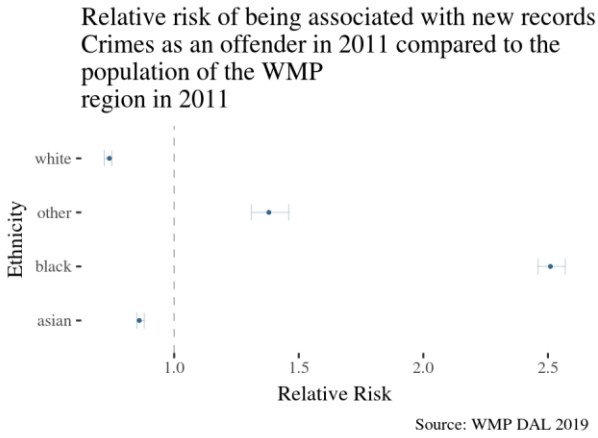
- The victim age distribution has a larger variance than the offender age distribution.
- Both distributions peak around the early to mid-twenties.
- There is an almost constant number of victims below the age of roughly 12 years which is not present in the offender distribution.

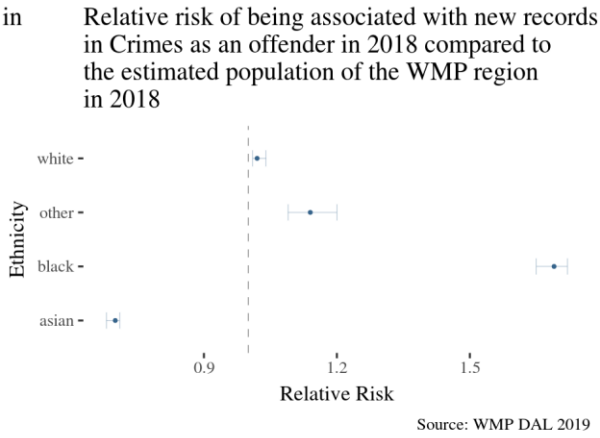
Distribution of age of nominals associated with new records in Crimes for Victims and Offenders from 2011 to 2018 inclusive

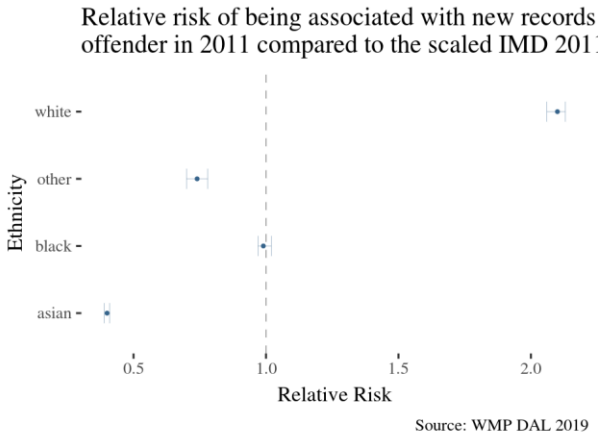


The remaining plots in this section show the estimates of relative risk.

The below plot shows the estimate of relative risk of being associated with a record in crimes as an offender for each ethnicity.

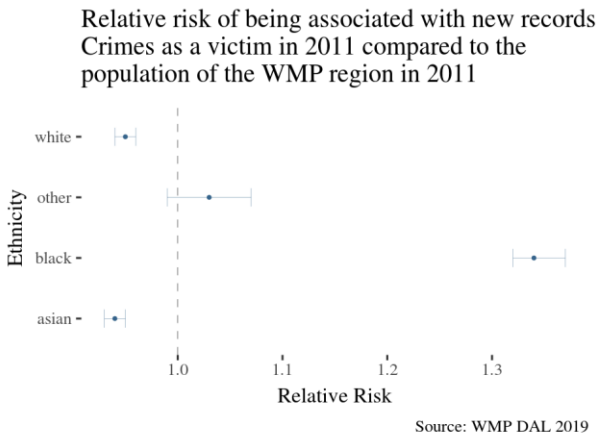


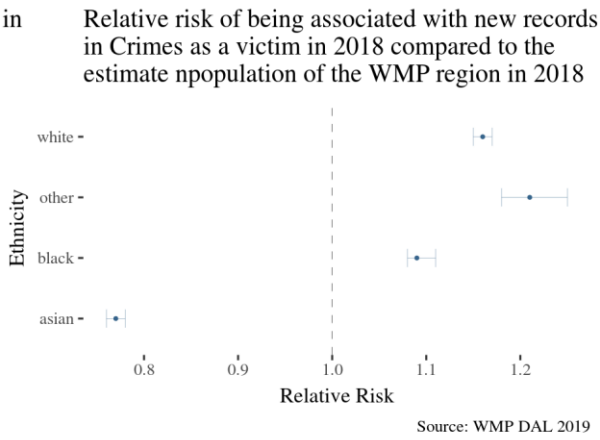


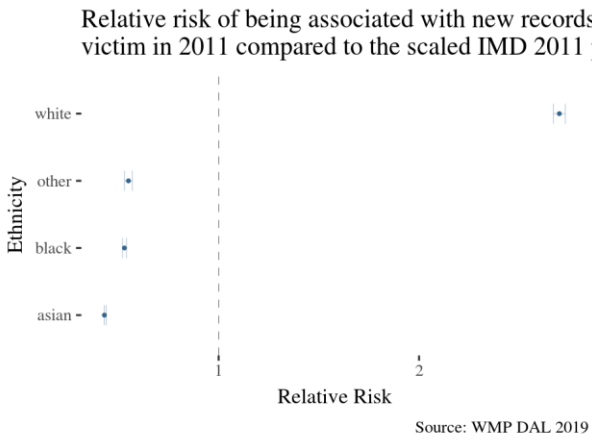


- These plots show the sensitivity of the relative risk to the population concerned. Using the base populations for 2011 or 2018 show, for example, a higher relative risk of being within the crimes system for black nominals compared to white nominals. However this difference reverses when the characteristics of locations of relatively high demand are taken into account.
- The relative risk of asian nominals being associated with new records in Crimes as an offender is lower in all three estimates, but by varying degrees.

The below plot shows the estimate of relative risk of being associated to a record in crimes as a victim for each ethnicity.





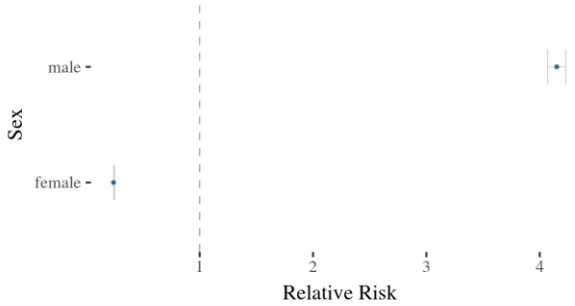


Similarly to above,

- The relative risk of white and black nominals being associated with new records in Crimes as a victim varies depending upon which population is used.
- The relative risk of asian nominals being associated with new records in Crimes as a victim is lower in all three estimates, but by varying degrees.

The following plot shows the estimate of relative risk of being associated to a record in crimes as an offender for each sex.

Relative risk of being associated with new records in Crimes as an offender in 2011 compared to the population of the WMP region in 2011

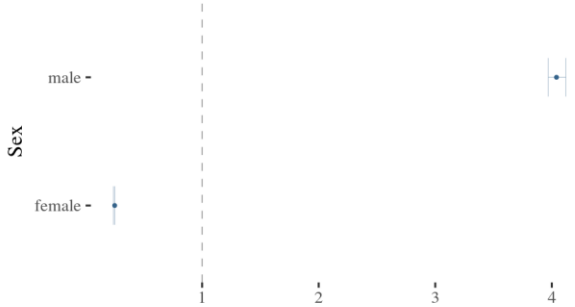


Source: WMP DAL 2019

Relative risk of being associated with new records in Crimes as an offender in 2018 compared to the estimated population of the WMP region in 2018

Source: WMP DAL 2019

Relative risk of being associated with new records in Crimes as an offender in 2011 compared to the scaled IMD 2011 population

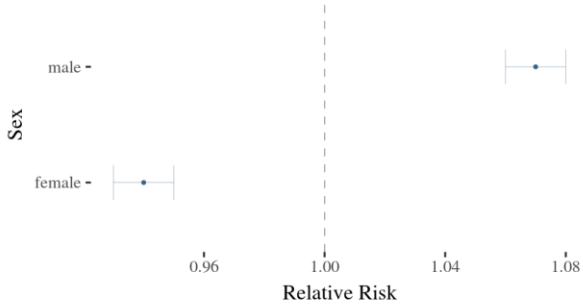


Source: WMP DAL 2019

From the above plots, males are estimated to be at least three times more likely to be associated with new records in Crimes as an offender compared to females.

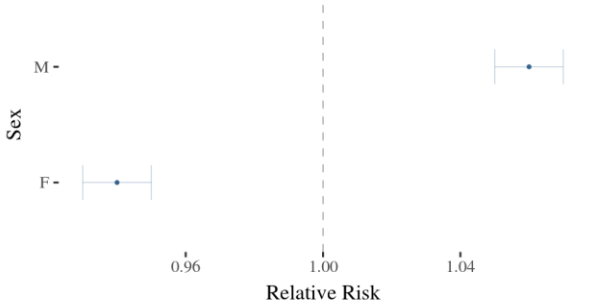
The below plot shows the estimate of relative risk of being associated to a record in crimes as a victim for each sex.

Relative risk of being associated with new records in Crimes as a victim in 2011 compared to the population of the WMP region in 2011



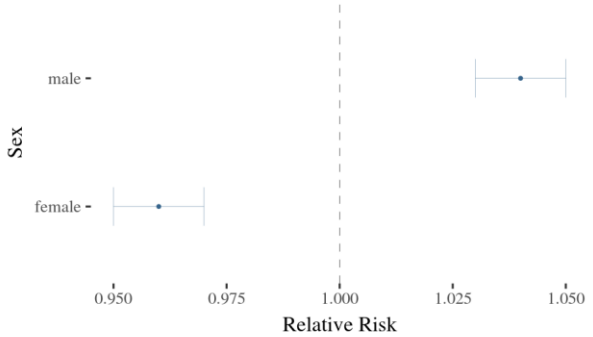
Source: WMP DAL 2019

Relative risk of being associated with new records in Crimes as a victim in 2018 compared to the estimated population of the WMP region in 2018



Source: WMP DAL 2019

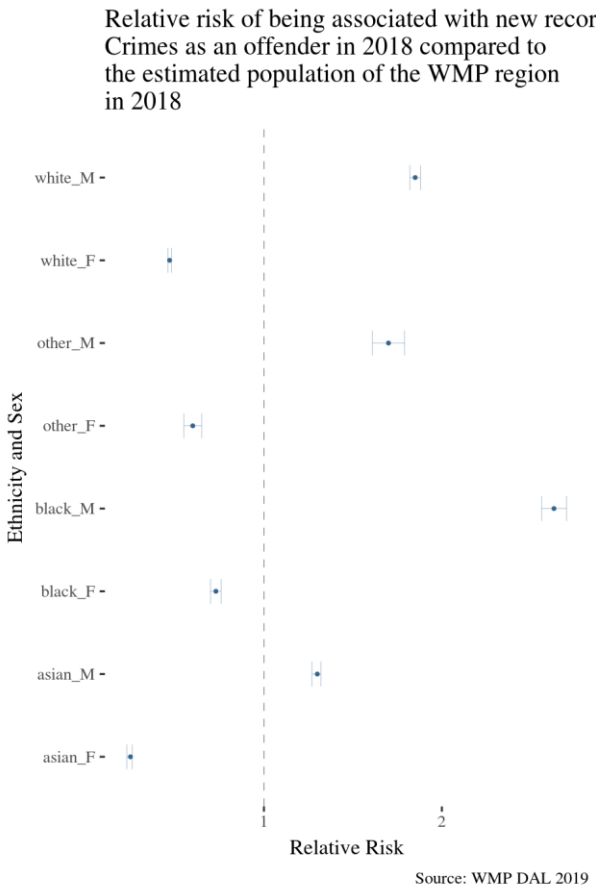
Relative risk of associated with new records in Crimes as a victim in 2011 compared to the scaled IMD 2011 population

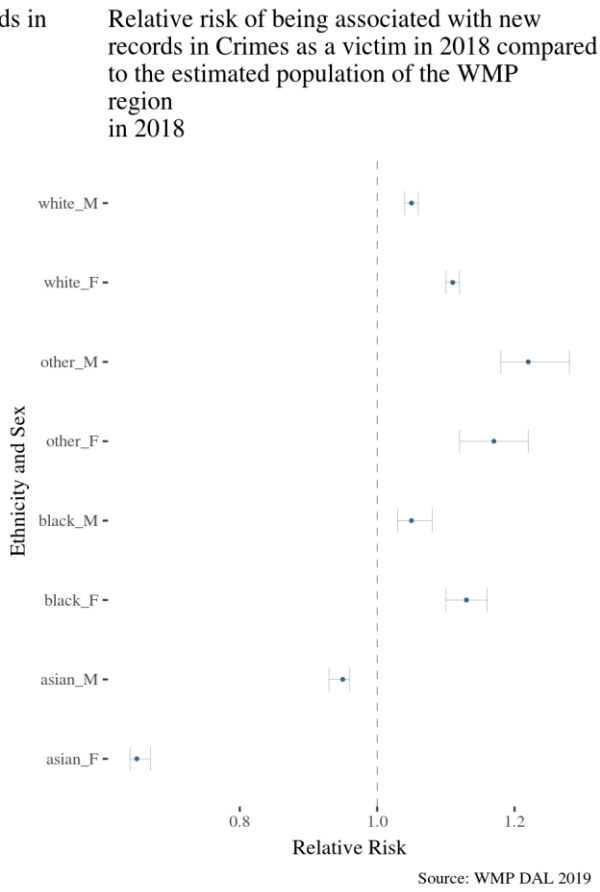


Source: WMP DAL 2019

There is little difference between the relative risk of males and females being associated with new records in Crimes as victims.

The below plot shows the estimate of relative risk of being associated to a record in crimes as an offender or victim for each ethnicity and sex combination.

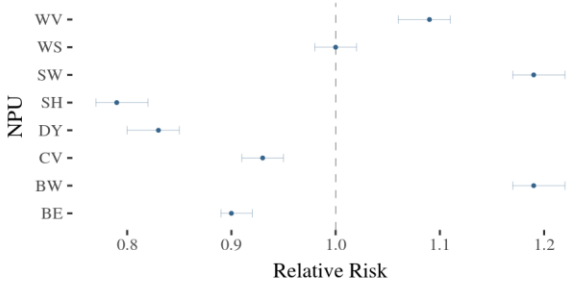




For the analysis of the Crimes system, there is no estimate of combined ethnicity and sex relative risk for the general 2011 population because the 2011 Census data does not provide the combined ethnicity and sex populations.

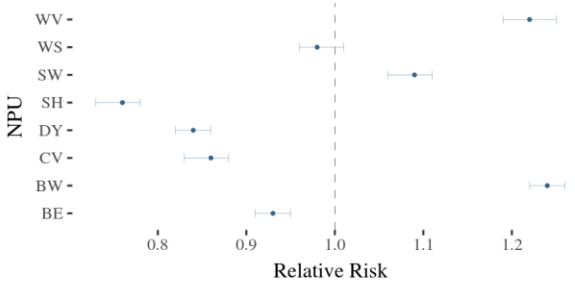
The below grid shows the estimate of relative risk of being associated to a record in crimes as an offender or victim in each NPU.

Relative risk of being associated with new records in Crimes as committing an offence in 2011 in a NPU compared to the population of the NPU in 2011



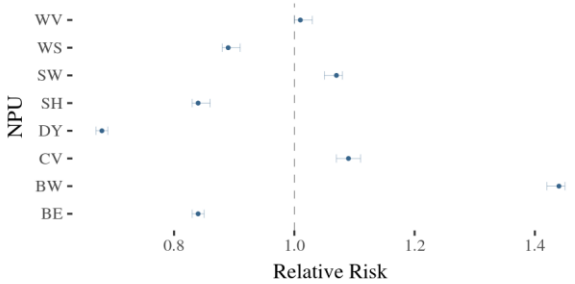
Source: WMP DAL 2019

Relative risk of being associated with new records in Crimes as an offender in 2018 compared to the estimated population of the WMP region in 2018



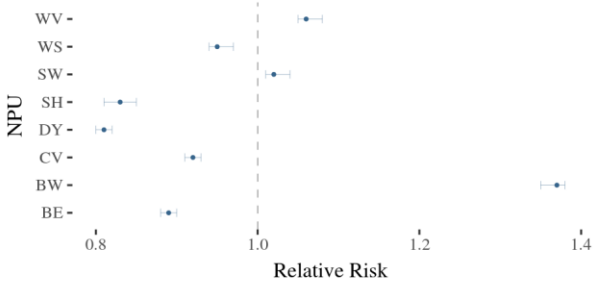
Source: WMP DAL 2019

Relative risk of associated with new records in Crimes as the victim of a crime in a NPU in 2011 compared to the population of the NPU in 2011



Source: WMP DAL 2019

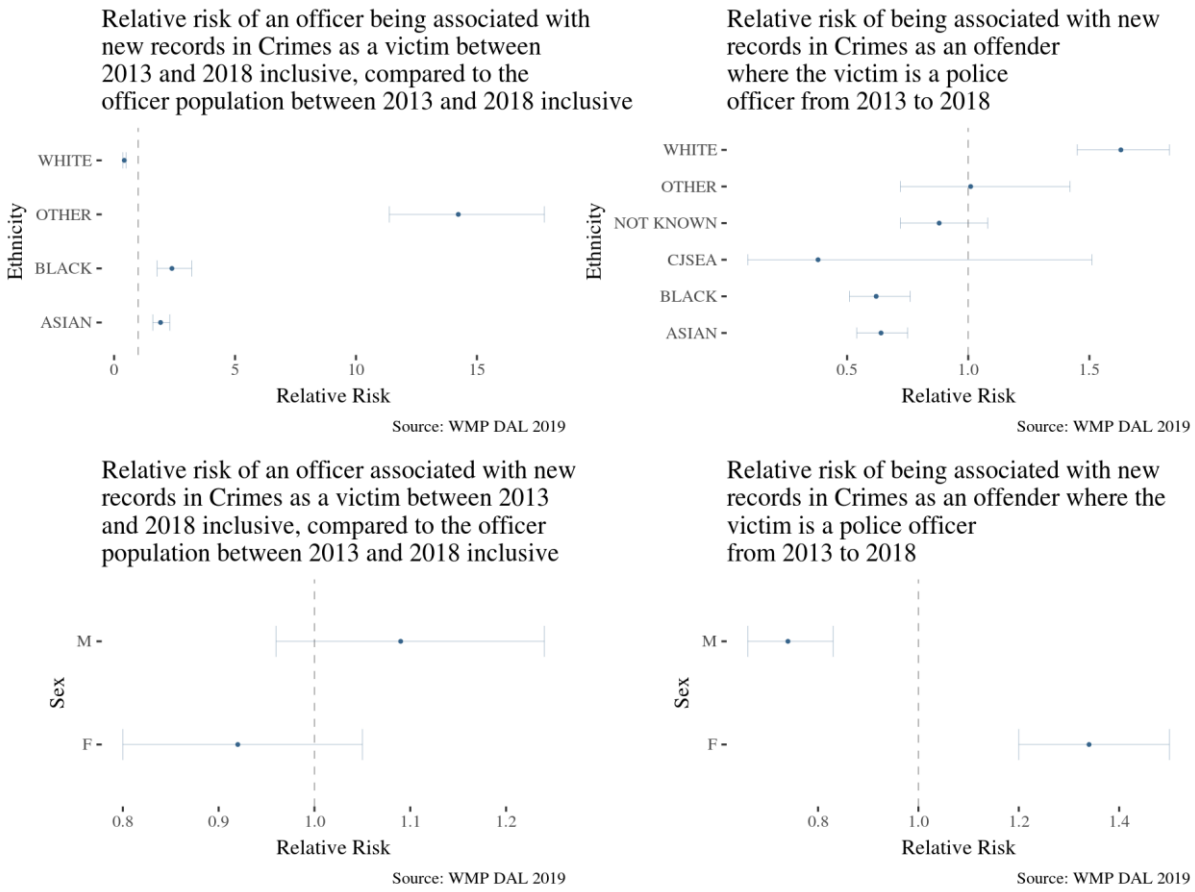
Relative risk of being associated with new records in Crimes as a victim in 2018 compared to the estimated population of the WMP region in 2018



Source: WMP DAL 2019

- There is a higher relative risk of nominals offending in the Birmingham West NPU (BW) compared to its population and a lower relative risk for the Solihull NPU (SH) compared to its population.
- There is a higher relative risk of nominals being victims in the Birmingham West NPU compared to its population and a lower relative risk for the Dudley NPU compared to its population.

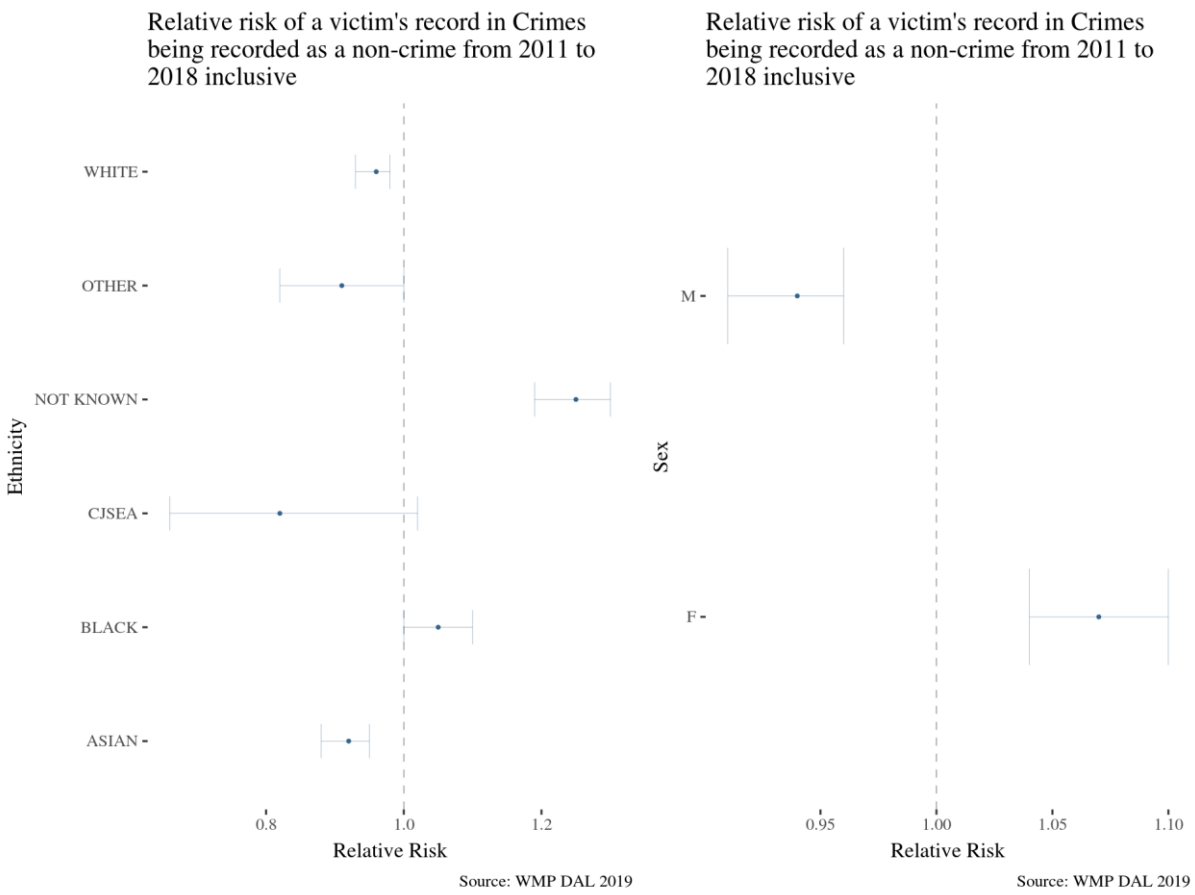
The following plot shows the estimates of relative risk where officers are the victim.



From the above plot:

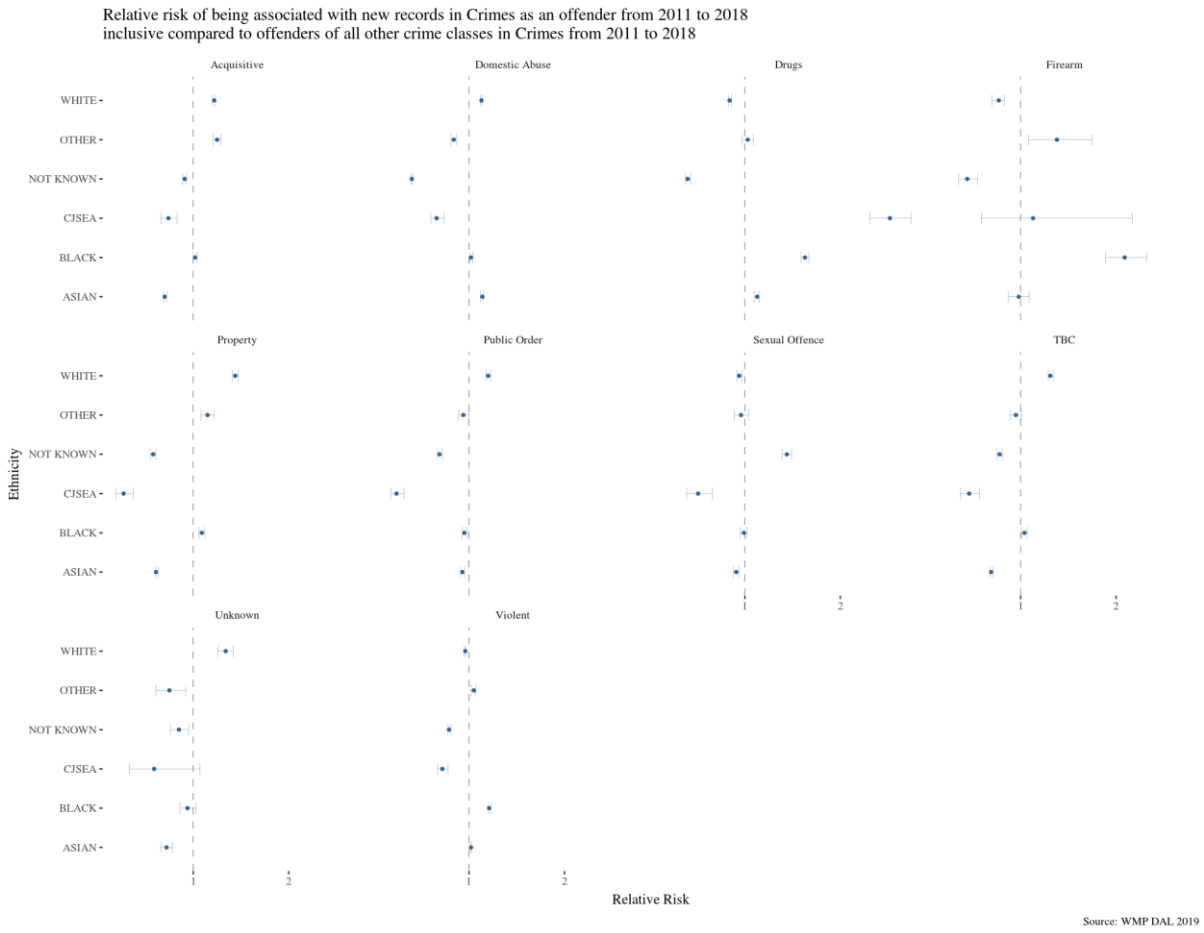
- Officers of other ethnicity have a higher relative risk of being associated with a record in Crimes as a victim of a crime. Black and Asian officers also have a higher relative risk than White officers.
- White nominals have a higher relative risk of being associated with a record in Crimes as an offender of a crime where an officer is the victim.
- Female nominals have a higher relative risk of being associated with a record in Crimes as an offender of a crime where an officer is the victim.

The below plot shows the estimate of relative risk of a record in crimes being recorded as a non-crime for the victims ethnicity and sex.



There is little difference in the relative risk of a record in Crimes being recorded as a "non-crime" compared to the victim's ethnicity or sex.

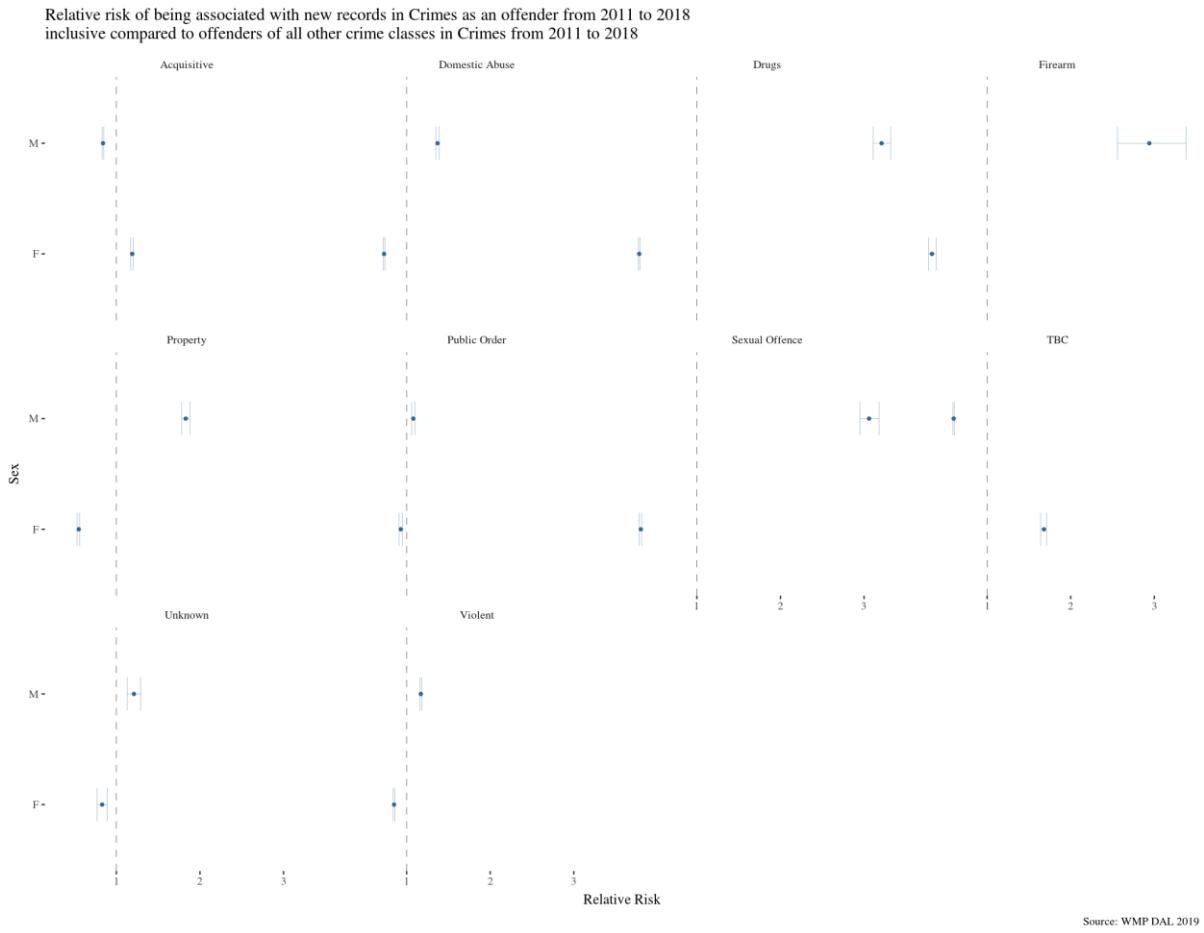
The following plot shows the estimates of relative risk for the most frequent crime classes for each ethnicity.



Some estimates that differ are:

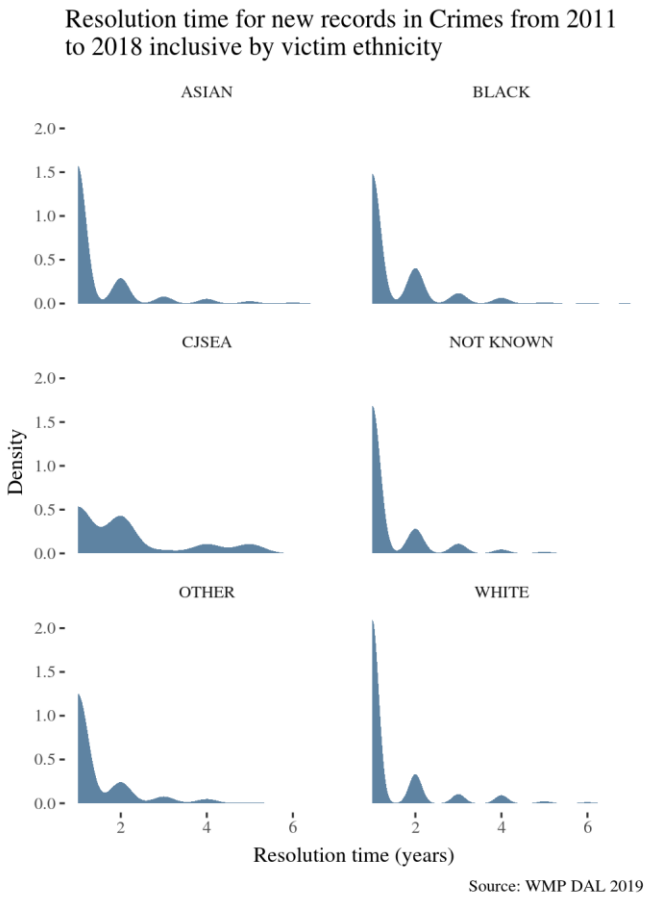
- There is a higher relative risk of CJSEA and black nominals being associated with new drugs crimes as offenders from 2011 to 2018.
- There is a higher relative risk of black nominals being associated with new firearm crimes as offenders from 2011 to 2018.
- There is a higher relative risk of white nominals being associated with new property crimes as offenders from 2011 to 2018 and a lower relative risk of CJSEA nominals being associated with new property crimes as offenders from 2011 to 2018.
- There is a higher relative risk of CJSEA nominals being associated with new public order crimes as offenders from 2011 to 2018.

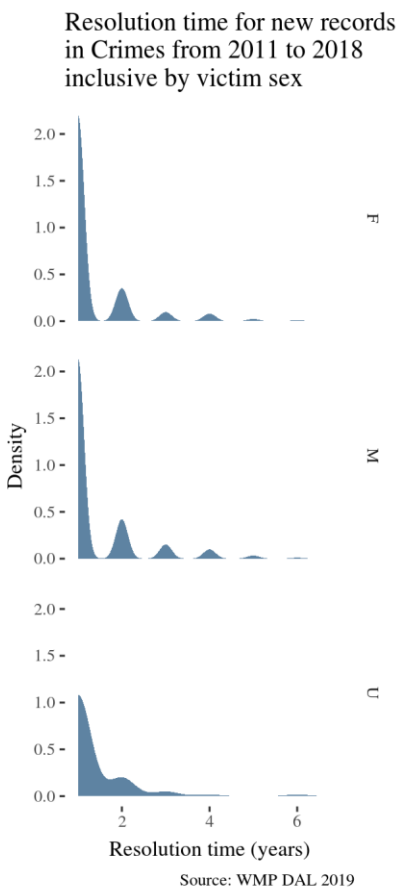
The following plot shows the estimates of relative risk for the most frequent crime classes for each sex.



There is a higher relative risk of male nominals being associated with new drugs, firearm, property and sexual offence crimes as offenders from 2011 to 2018 compared to females.

The final plot shows the time to resolution for crimes by victim ethnicity.





There is little difference between the distributions of resolution times.

5 Custody

The ICIS database contains all details relating to nominals who have entered custody, this has been queried to obtain details of the nominals, date of birth, ethnicity and sex, and their offence(s) including its severity[4].

5.1 Data

After a series of joins, the fields returned from the ICIS database are:

Field	Data Type	Description
crime_crime_ref	int	Crime reference (same as in Crimes)
nominal_ref	character	Nominal reference
offence_number	character	Offence reference
offence_title	character	Offence description
role_type	character	Role type in crime
custody_crime_ref	character	Custody reference
offence_no	int	Cumulative count of offences by nominal
ethnicity_detail	character	Ethnic description
age	int	Age of offender (years)
sex	character	Sex of offender
occupation	character	Occupation of offender
ethnicity	character	Ethnic group of offender
severity	double	Crime severity

Data dictionary of ICIS data returned

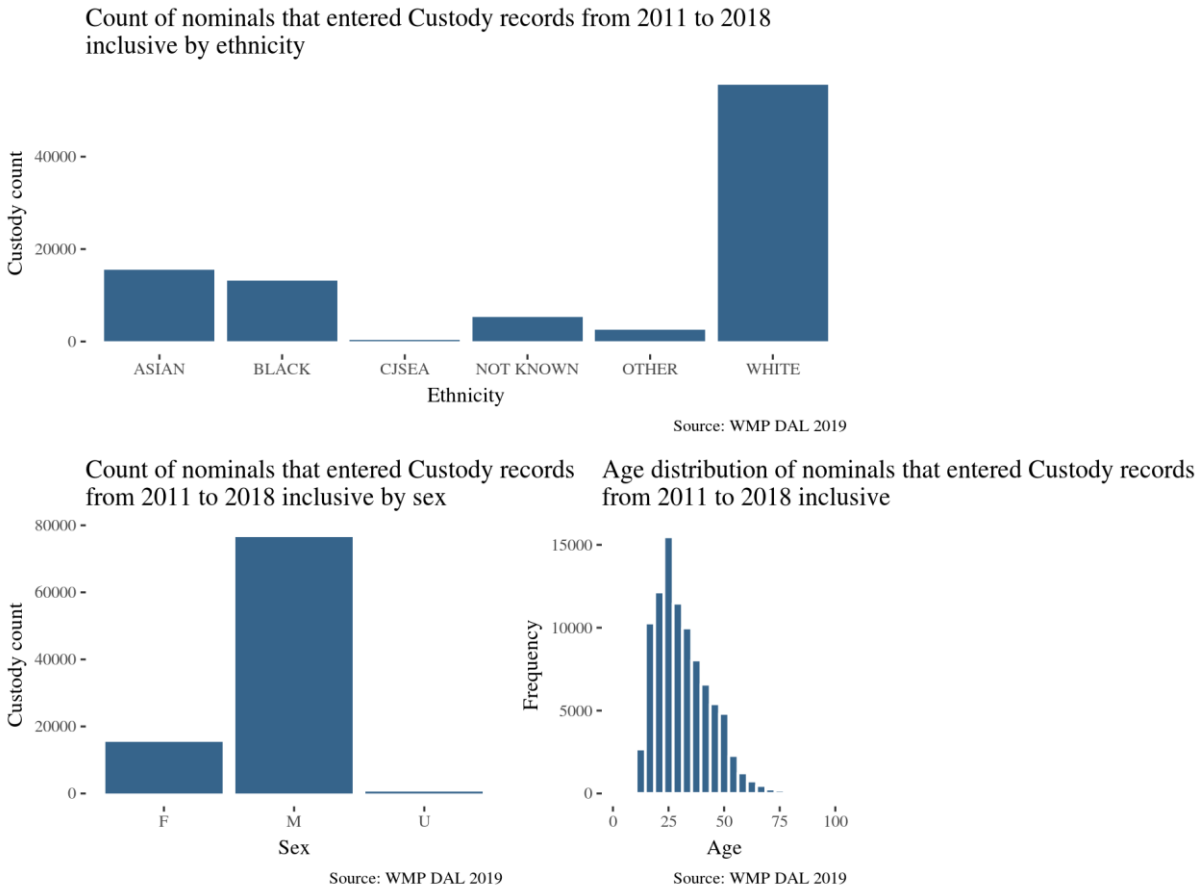
5.2 Results

Again, a series of contingency tables are constructed for comparison with the "population" data across a characteristic. Here there is less debate about what is considered the "population", since only nominals that have committed an offence can be taken into custody, therefore the "population" data for comparison with the custody records is the records of offenders of crimes.

The following plots summarise the analysis.

[4] The crime severity is calculated as the average of the normalised Cambridge Harm Index and the normalised ONS Crime Severity score (i.e. both measures are normalised for each crime for each nominal so that they are on the scale).

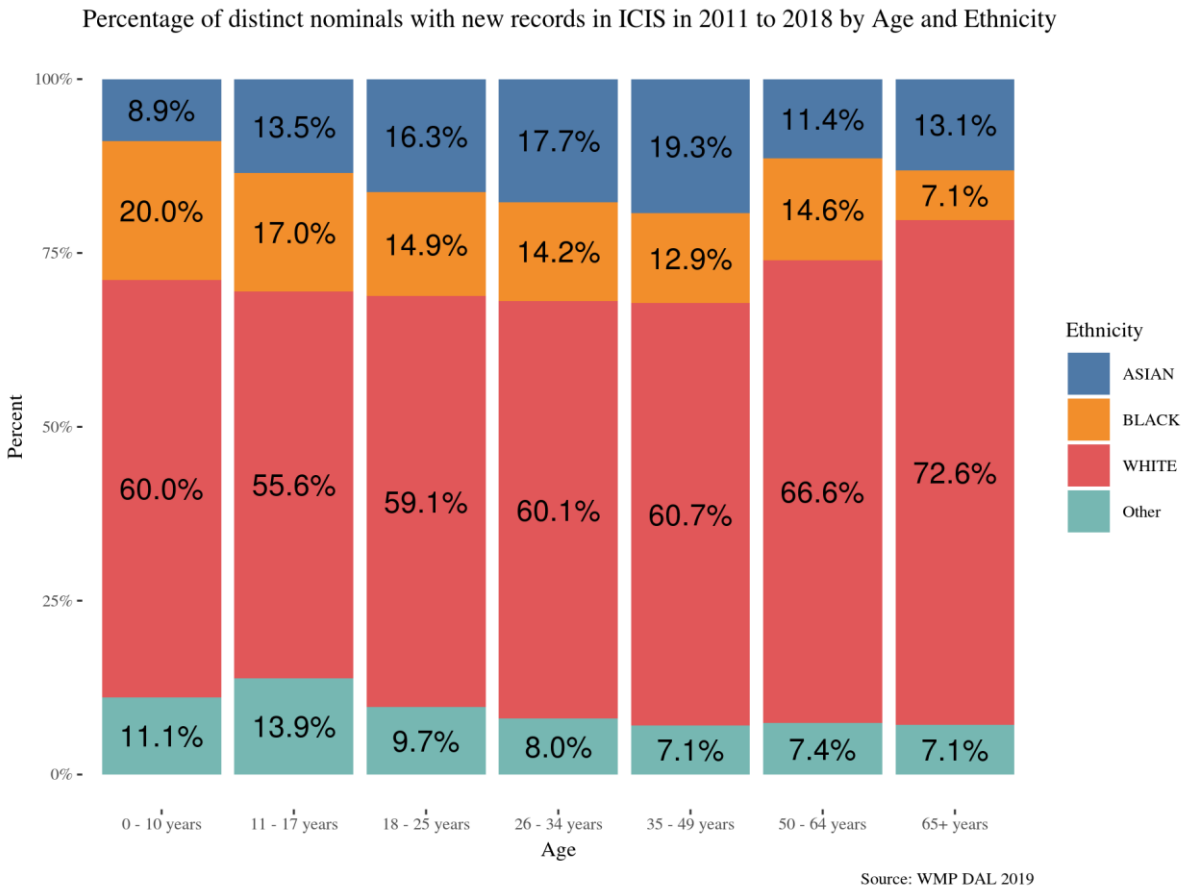
The first plot is a grid containing three plots. In the top half is the count of distinct nominals of each ethnicity, the bottom left quadrant is the count of distinct nominals per sex and the bottom right quadrant is the distribution of distinct nominal age.



From the above plot:

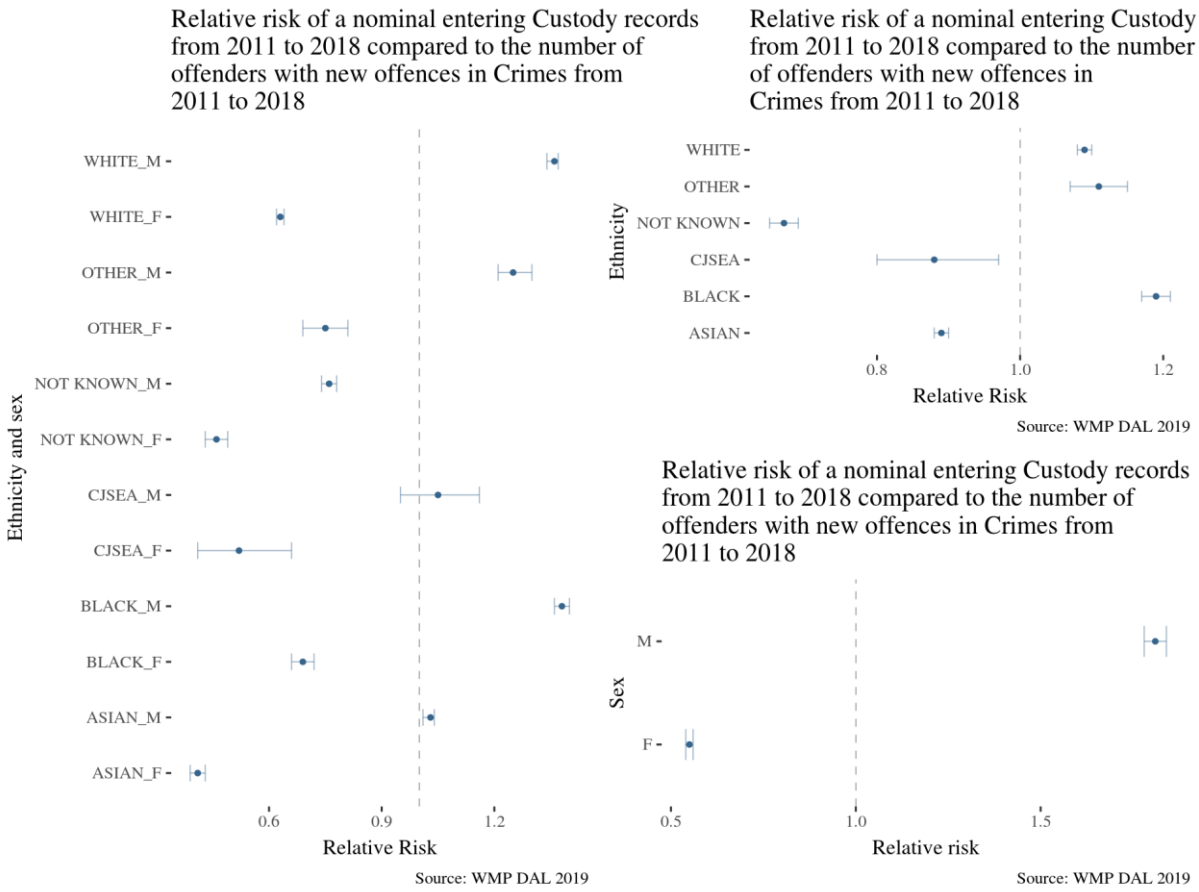
- There appears to be an increase in the number of black nominals compared to Crimes.
- There is a smaller proportion of records in ICIS with sex equal to “F” (female) compared to the number of offender records in Crimes with sex equal to “F” (female).
- The age distribution is very similar to the age distribution of offenders in Crimes.

The second plot shows the percentage of distinct nominals of asian, black, white and other ethnic groups per age group.



- The percentage of white nominals generally increases as the age increases.
- The percentage of black nominals is relatively constant (although does reduce between the age groups).

The final plots show all estimates of relative risk.



To summarise the above plots:

- Male nominals have a larger relative risk of entering custody between 2011 and 2018.
- Black nominals have a larger relative risk of entering custody between 2011 and 2018.
- Black and white males have larger relative risk while asian females have a smaller relative risk of entering custody between 2011 and 2018.

6 Stop and Search

The Stop and Search database contains the details of all Stop and Search incidents including details relating to the nominal and the record. An officer could Stop and Search someone if they have "reasonable grounds" to suspect a nominal is carrying something they shouldn't or without "reasonable grounds" if it is approved by a senior officer and they are acting suspiciously or in a specific location or area (gov.uk 2011). The database has been queried to obtain information relating to the nominal as well as details relating to the Stop and Search.

6.1 Data

After some feature engineering, the fields returned from the Stop and Search database are:

Field	Data Type	Description
stop_date	datetime	Date and time of Stop and Search
record_standard	character	Whether the record meets the standard
ethnicity_detail	character	Nominal ethnic description
sex	character	Nominal sex
dob	date	Nominal date of birth
outcome	character	Outcome of Stop and Search
age	int	Age of nominal at Stop and Search (years)
ethnicity	character	Ethnic group of nominal
NPU	character	NPU of Stop and Search
geometry	point	Easting and northing point of Stop and Search

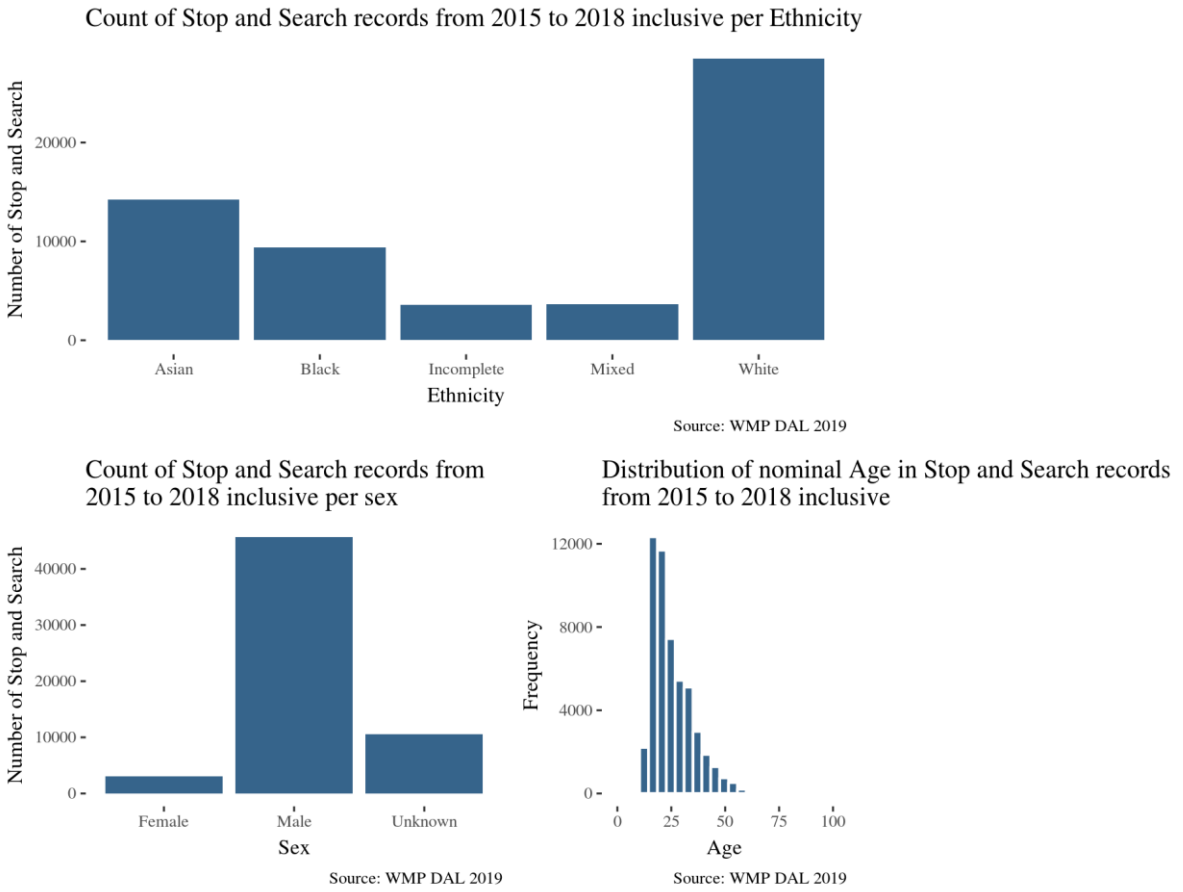
Data dictionary of Stop and Search data returned

6.2 Results

Again, here there could be some debate about what is considered the relevant "population" (for the denominators in the relative risk calculations), since anyone in a specific area or anyone acting suspiciously can be searched (gov.uk 2011) this could include anyone in the population but it may be that Stops and Searches are predominantly undertaken in areas of higher crime rates which do not have populations that reflect the ethnic makeup of the population as a whole. It is also the case that some areas where Stops and Searches are undertaken have relatively little population compared to the number of stops undertaken (e.g. Birmingham City Centre and Birmingham Airport). Hence a comparison is made with the Crimes offenders and both the 2011 Census and the IMD population.

The following plots summarise the analysis.

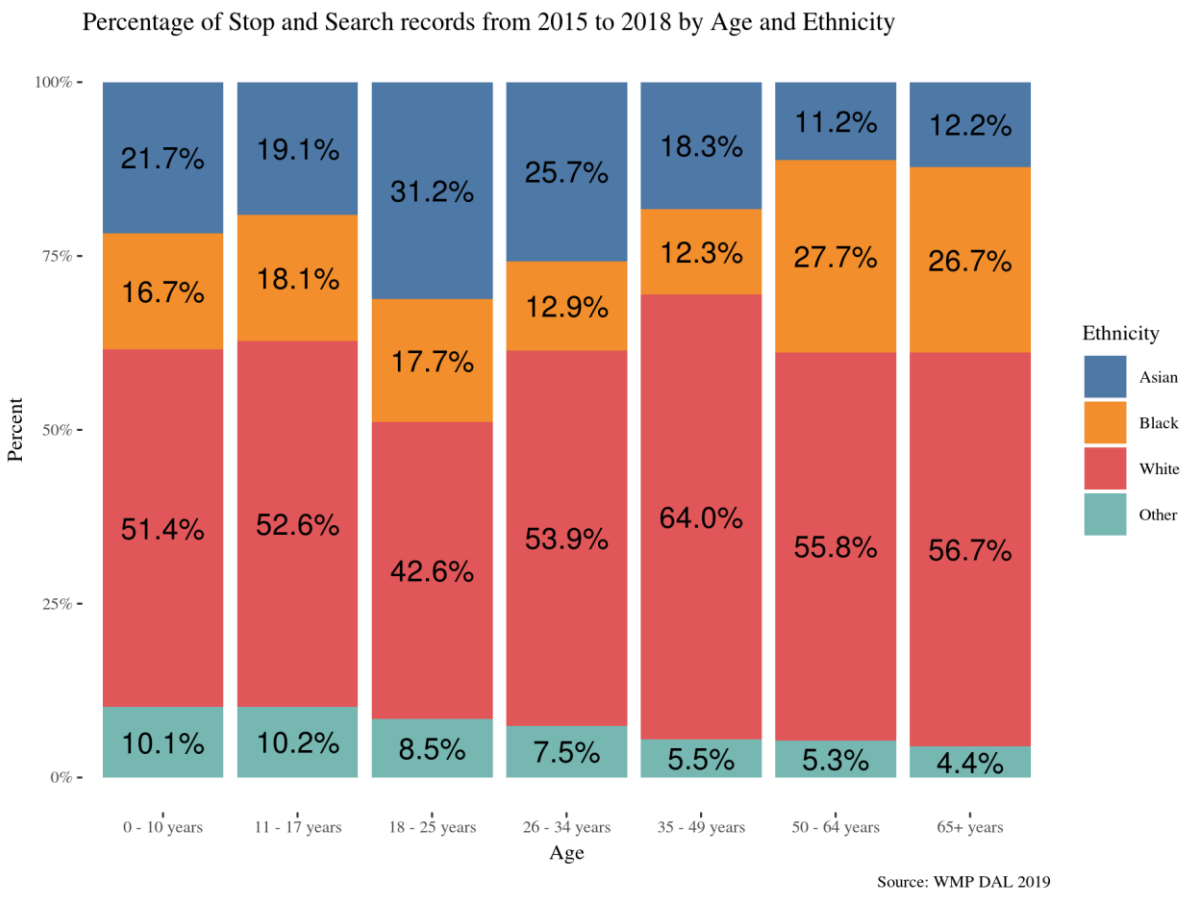
The first plot is a grid containing three plots. In the top half is the count records of each ethnicity, the bottom left quadrant is the count of distinct nominals per sex and the bottom right quadrant is the distribution of distinct nominal age.



From the above plot:

- There is a larger proportion of asian and black nominals compared to both the Crimes population and the 2011 Census.
- There is a larger proportion of male nominals compared to the 2011 Census and the Crimes population.
- The age distribution is skewed towards younger nominals compared to Crimes offenders, which is most likely due to the Crimes system including many crimes to which Stop and Search is not addressed (e.g. domestic abuse).

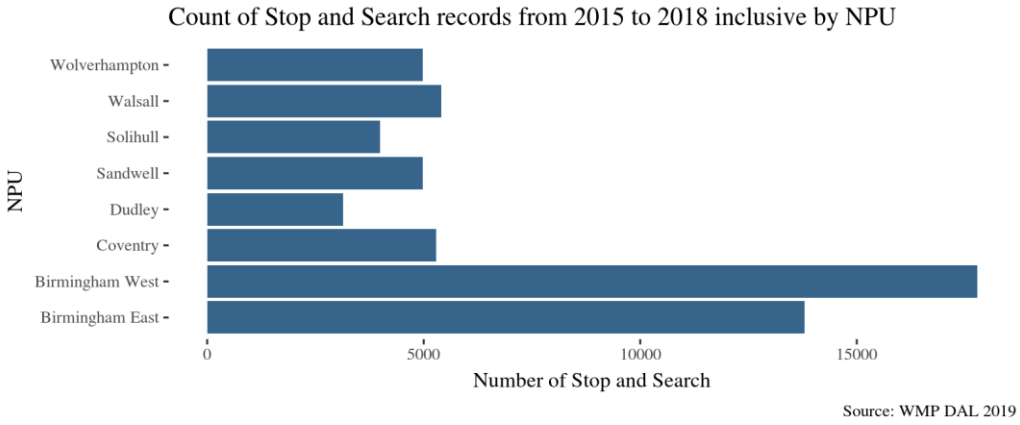
The second plot shows the percentage of distinct nominals of asian, black, white and other ethnic groups per age group.

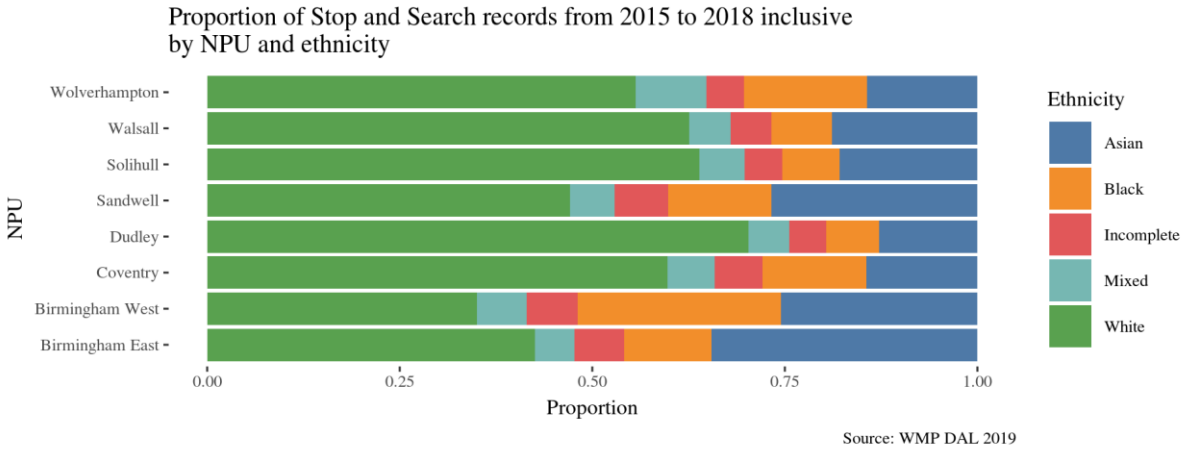


Some observations from the above plot are:

- The percentage of white nominals is generally less compared to the equivalent plot for the Crimes population and there is no great increase as the age increases.
- The percentage of asian and black nominals is larger in each age group compared to the equivalent plot for the Crimes population.

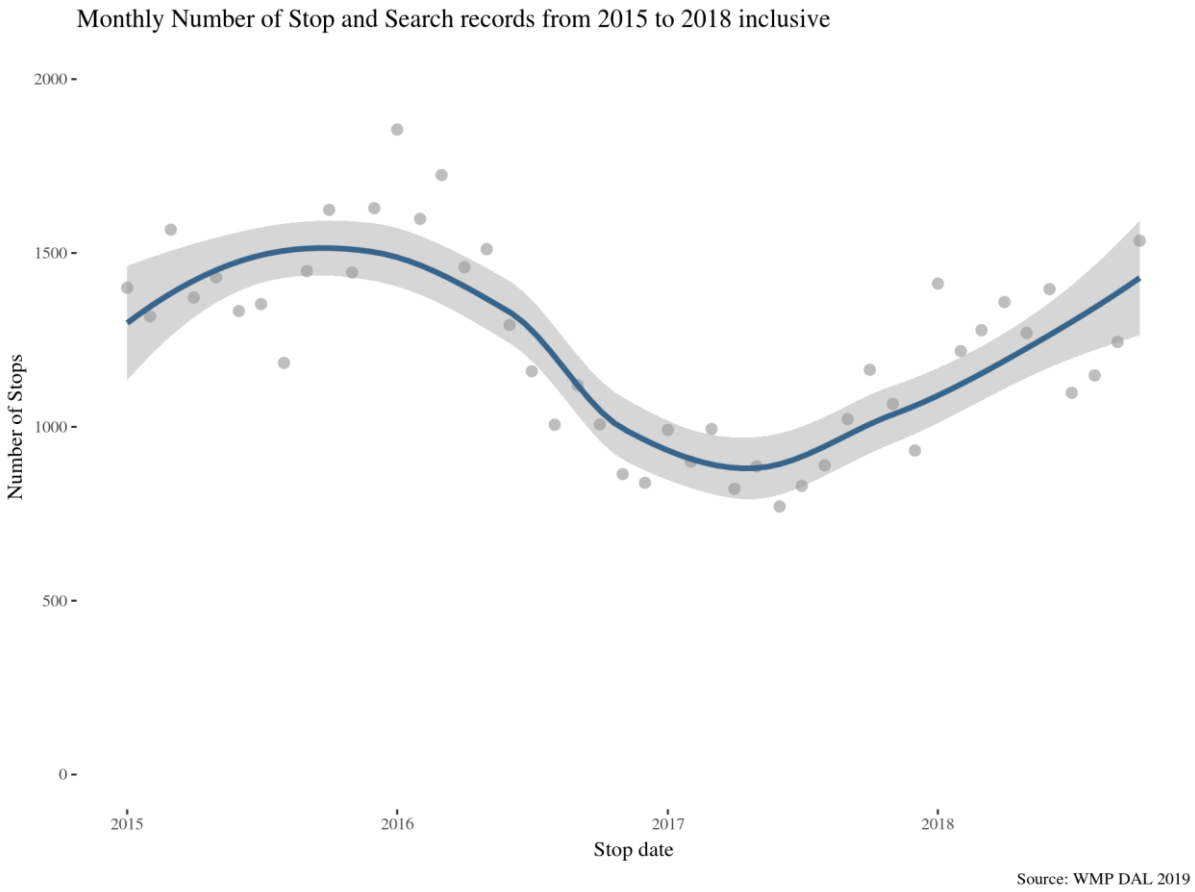
The next plot is split in two, the top half shows the number of Stop and Search records per NPU, while the bottom half shows the proportion of Stop and Search records of each subject ethnicity per NPU.





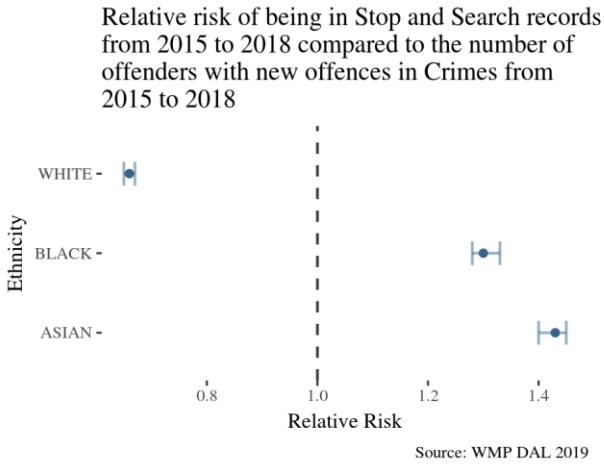
Birmingham East and Birmingham West have the two largest numbers of Stop and Searches and the lowest proportion of white subjects; this could be due to the ethnic make-up of these NPUs.

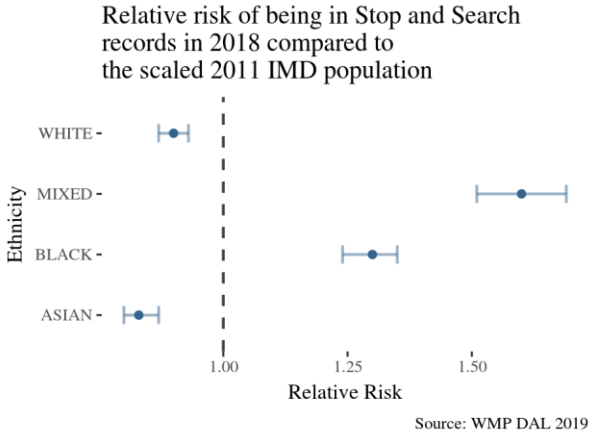
The following plot shows the monthly number of Stop and Search records from 2015 to 2018.

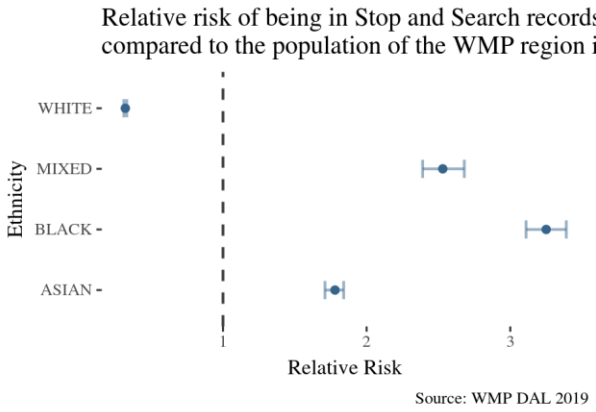


From the start of 2016 to the start of 2017 the number of monthly Stop and Searched decreased by roughly a third. This has since increased to figures similar to before the decrease.

The following plot shows the relative risk of each ethnicity in the Stop and Search records.





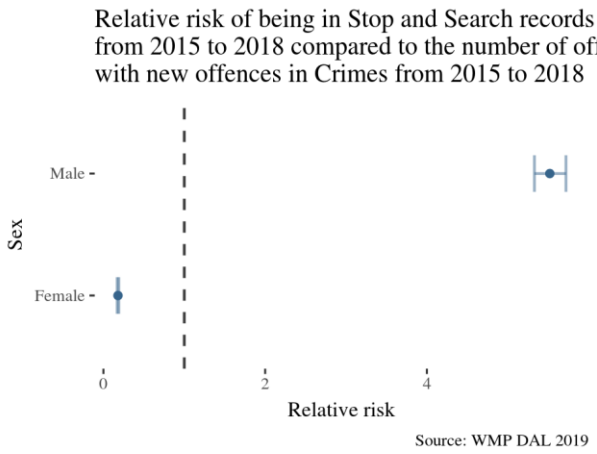


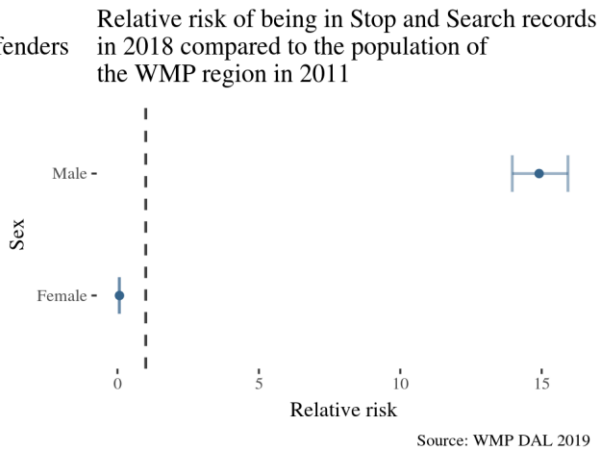
Some observations from the above plots are:

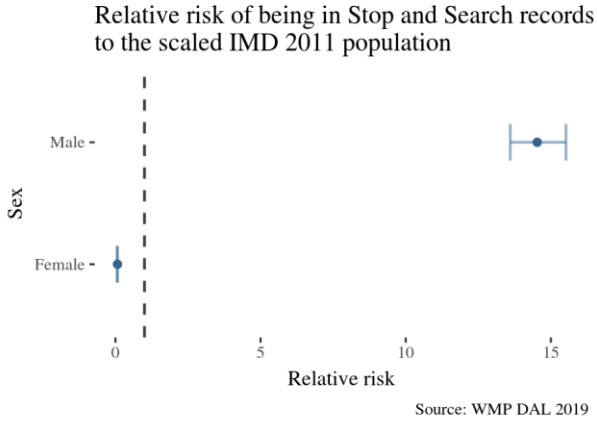
- Black and asian nominals have a larger relative risk of being in the Stop and Search records while white nominals have a smaller relative risk of being in the Stop and Search records, when compared to the Crimes offenders population[5].
- Black and mixed nominals have a larger relative risk of being in the Stop and Search records when compared to the scaled IMD population.
- All ethnicities except for white, have a larger relative risk of being in Stop and Search records compared to the 2011 Census population.
- It is of note that Stops and Searches of black nominals using the Crimes and scaled IMD populations, whilst still higher compared to whites does not exhibit as large a difference when compared to the (Census based) population as a whole (moving from a relative risk of circa 3.2 to circa 1.3). This indicates a locational effect whereby the ethnic makeup of the areas where Stops and Searches are predominantly undertaken is different from that of the population as a whole.

The next plot shows the relative risk of each sex in the Stop and Search records.

[5] Crimes does not use the ethnicity label "MIXED" hence there is no estimate of relative risk for that ethnic group when compared to the Crimes offenders population.

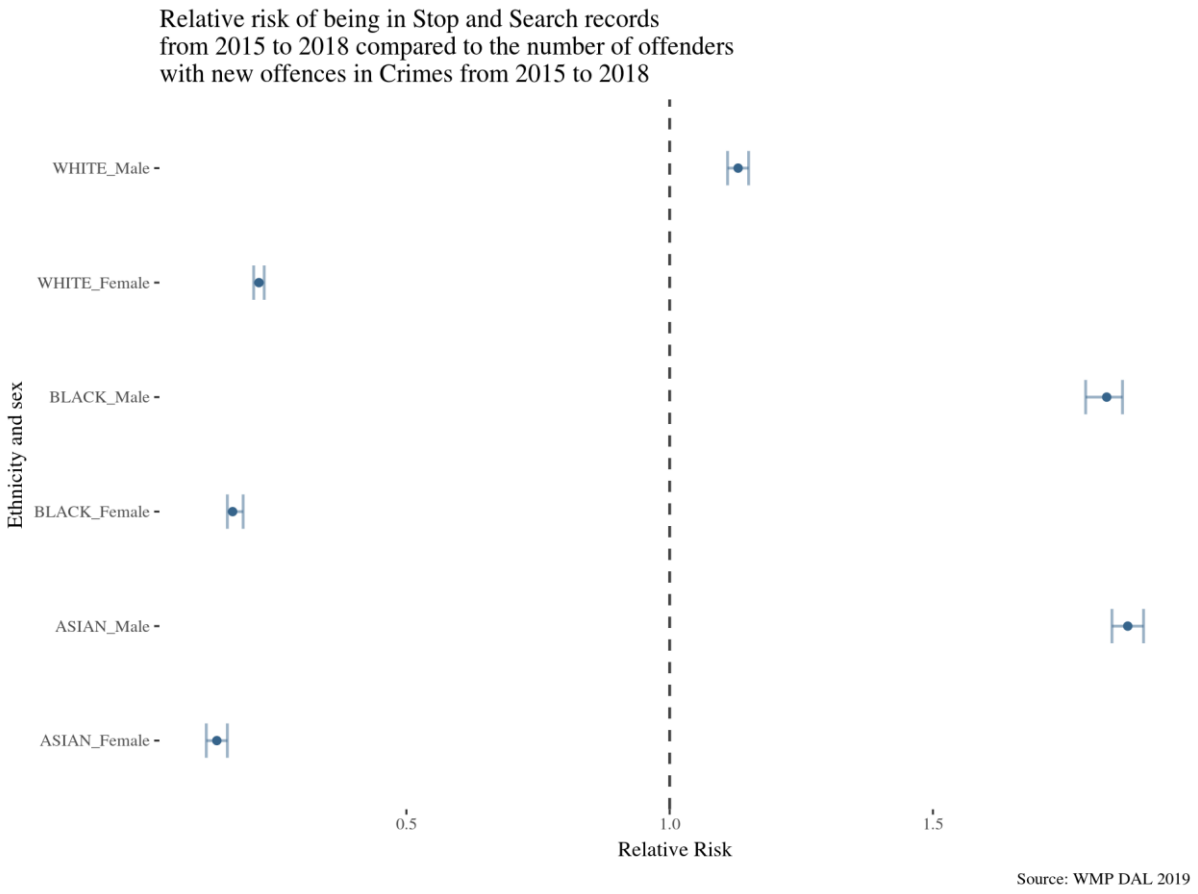






In all three comparisons, male nominals have a much larger relative risk of being in the Stop and Search records compared to female nominals. The 2011 Census population and IMD relative risks are very similar - this is due to the similarity of male / female split for the 2011 Census and the scaled IMD population - however they are not the same.

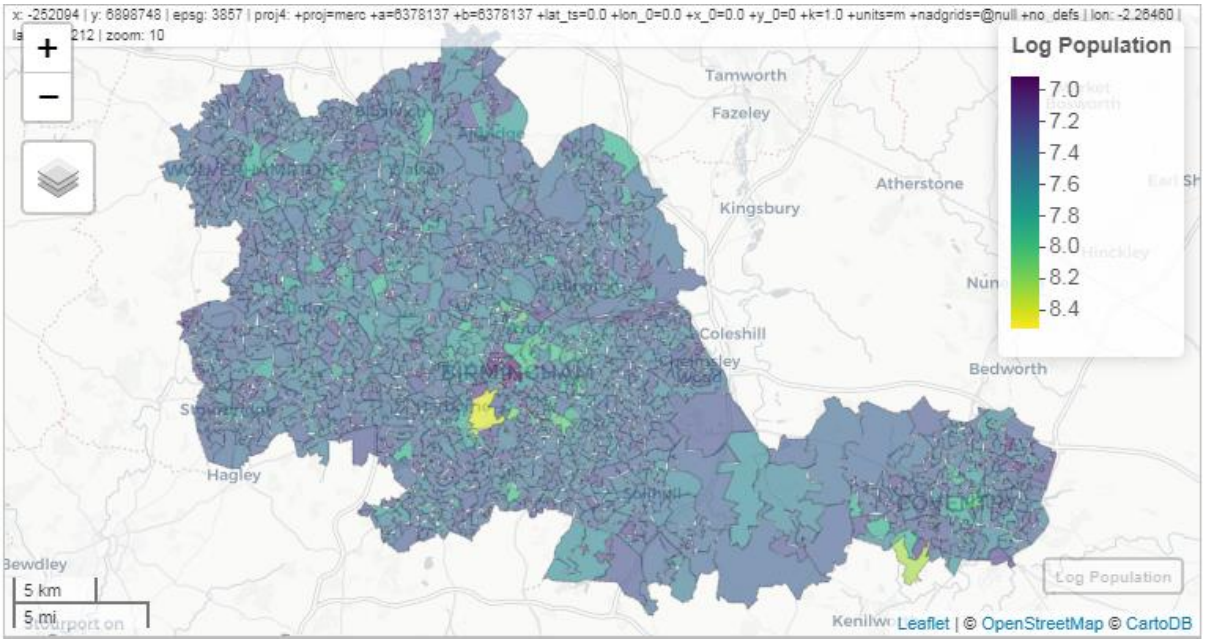
The final relative risk plot shows the relative risk of each ethnicity and sex combination in the Stop and Search records.



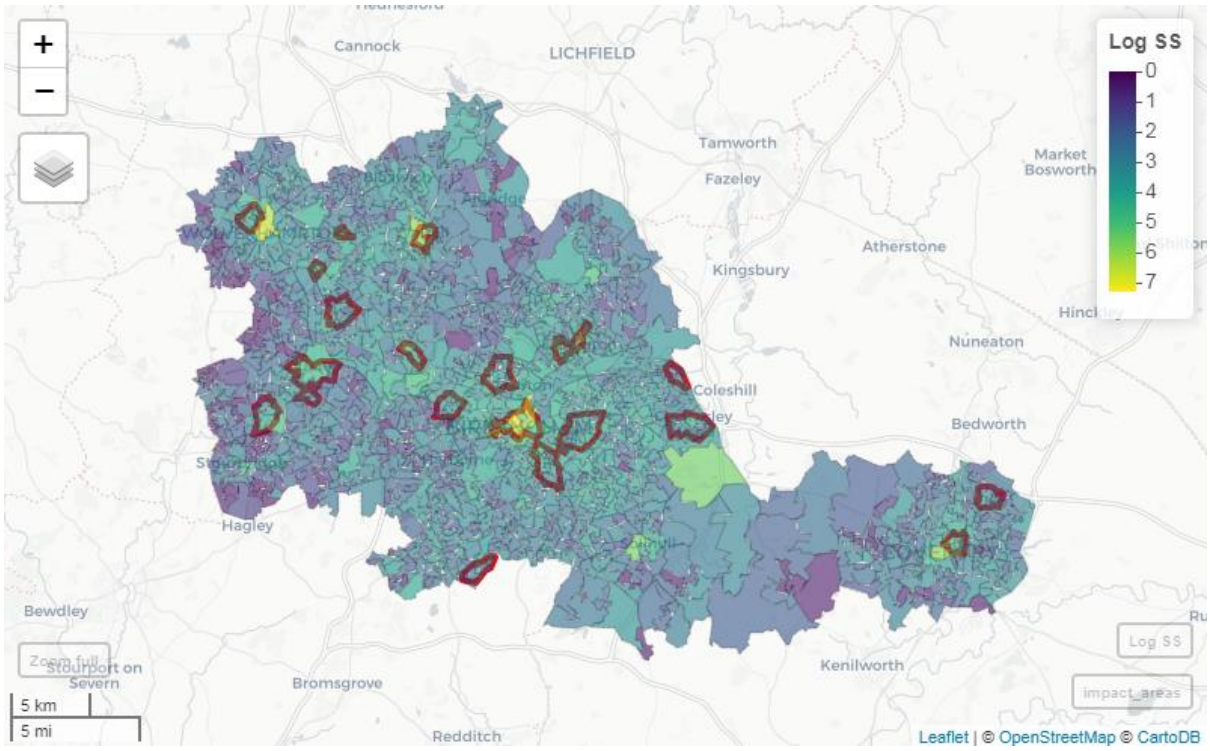
Black and asian male nominals have a larger relative risk of being in the Stop and Search records while white, black and asian females all have a smaller relative risk of being in the Stop and Search records.

The next two plots are maps of the WMP region, each LSOA is coloured on a continuous scale, yellow representing high values and purple representing low values. The former is coloured by population (the $\log_e$ of population) and the latter by the number of Stop and Search records (again, the $\log_e$ of the number of Stop and Search records) along with WMP 2019 impact areas (highlighted by a thick red border). The 19 impact areas highlight areas of long-term demand across several crime types and aim to reduce crime and harm in each impact area. Most of the impact areas cover areas where many Stop and Searches occur.

$$\ln(Population)$$



$\ln(\textit{Stops and Searches})$



LSOA **E01033620** is present in the previous two plots, this LSOA is in the very centre of Birmingham and has the highest number of Stop and Search records amongst the LSOAs. It also has a very small population (it is a dark shade of purple in the first map).

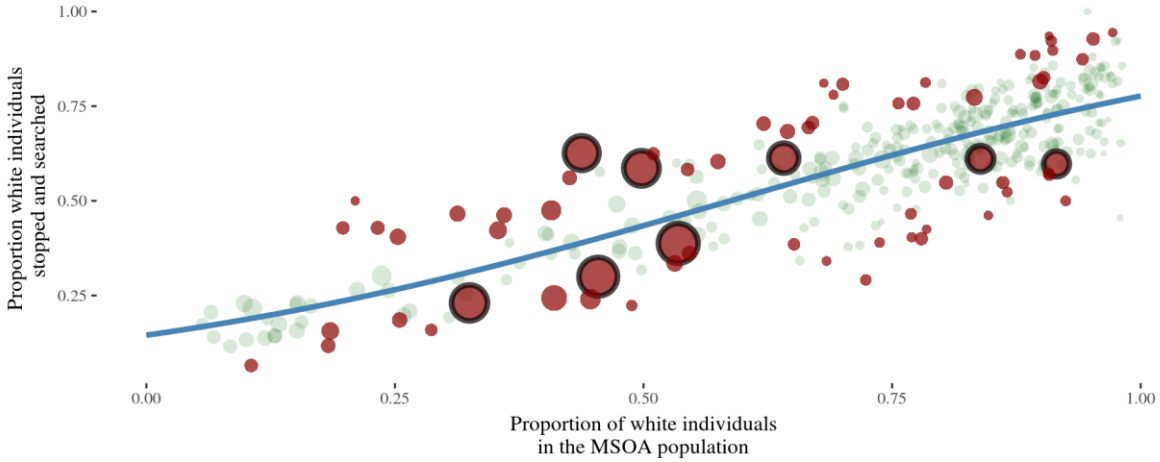
LSOA **E01010109** is again present in the previous two plots, it is the light green coloured LSOA in north Solihull with a high number of Stop and Searches and a small population. This LSOA contains Birmingham International Airport (where a large number of Stops and Searches take place).

Both of the LSOAs mentioned above have potentially high “traffic”, in other words, a lot of nominals travel to these LSOAs that do not reside there. This would account for the high Stop and Search rate compared to the low population. These LSOAs are evidence that comparing the number of Stop and Search records to the number of nominals residing in a LSOA is not necessarily the whole story.

LSOAs are small areas, having an average population of circa 1,500, this results in a large number of LSOAs with very few or many Stop and Searches. To mitigate this the LSOAs have been aggregated to MSOA (Middle layer Super Output Area) which has an average population of circa 7,500. The final plot shows the proportion of ethnicity residing in the MSOA on the horizontal axis against the proportion of each ethnicity stopped and searched in the MSOA. The blue line represents a binomial model of the ethnic proportion in a particular MSOA, accounting for the amount of evidence (number of total Stop and Searches) in a MSOA. The green points lie inside of the 99% prediction interval and the red ones lie outside, the points with a thick black border are MSOAs from city centres or the MSOA which contains Birmingham International Airport.

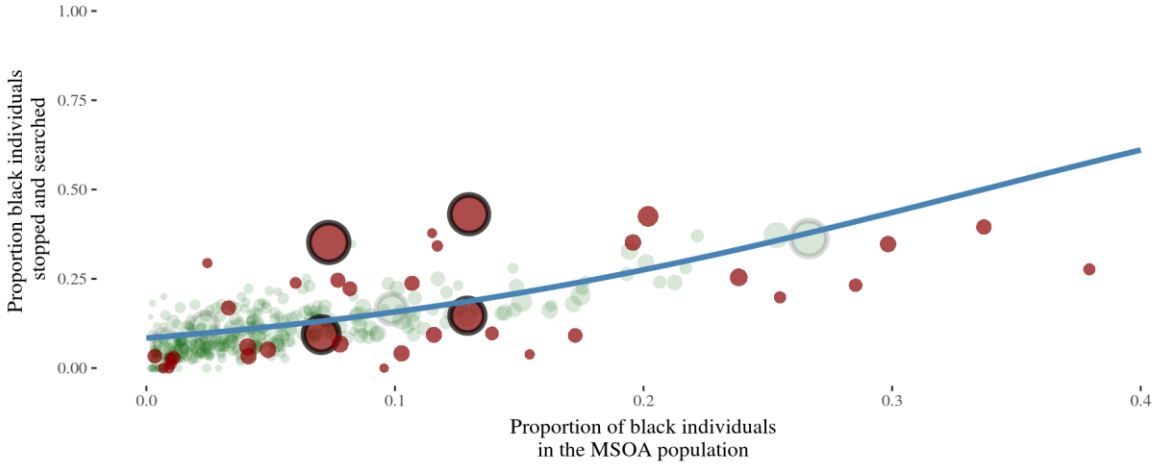
Stop and Search by MSOA for white nominals

Dark red points are LSOAs outside of a 99% credible prediction interval.
Points with a black border are city centre MSOAs or the MSOA which contains the airport.

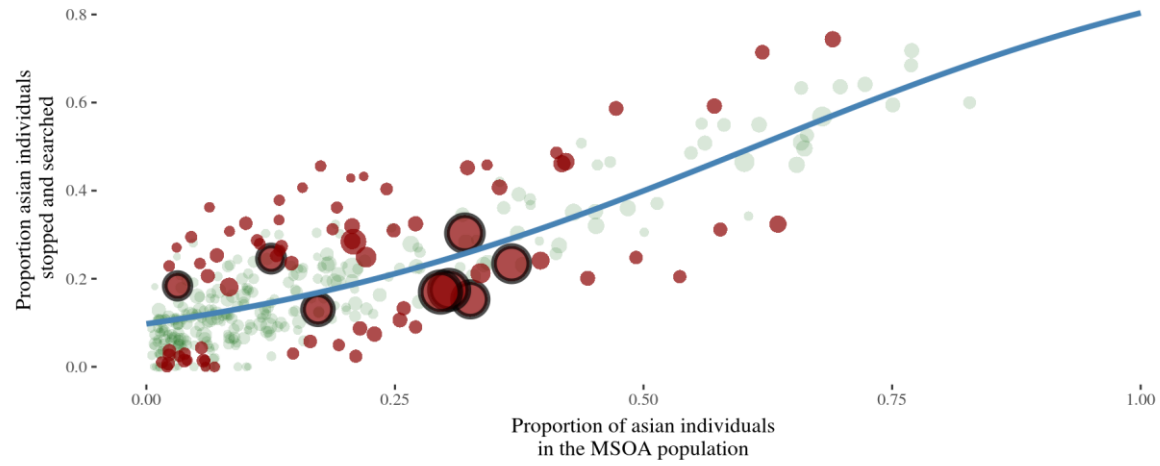


Stop and Search by MSOA for black nominals

Note that the horizontal axis has been limited to 0.4 because there are no MSOAs with a proportion of black residents greater than 0.4.



Stop and Search by MSOA for asian nominals



Source: WMP DAL 2019

The mean prediction shows a positive relationship between the proportion of nominals residing in a MSOA of a particular ethnicity and the proportion of nominals that are stopped and searched of the same ethnicity in that MSOA across all three ethnicities. The intercept of the blue line is non-zero for all plots, this shows that when there are no nominals of a particular ethnicity residing in an MSOA there are some stopped and searched, this indicates movement of nominals between the MSOAs and therefore the ethnic makeup of nominals stopped and searched in a particular MSOA is not only dependent on the ethnic makeup of residents of the MSOA. Note that this does not take into account potential clumping, in other words, each Stop and Search may not be independent of the next. If a group of nominals is stopped and searched are they more likely to be of the same ethnicity? Grouping the Stop and Search records together based on time of Stop and Search gives roughly 79% of Stop and Searches are on a single nominal, 14% are estimated to be on a pair of nominals and 5% are estimated to be on a group of three nominals.

7 Use of Force

WMP requires officers to keep a record as to whether force has been used and these records are the focus of this section. A Use of Force refers to a police officer using a forceful tactic on a nominal, this could include handcuffing while making an arrest through to pointing a firearm at a nominal. A Use of Force incident refers to one officer's Use of Force on a nominal, therefore if several officers or nominals are involved this would constitute multiple Use of Force incidents and result in multiple Use of Force records (Office 2018).

7.1 Data

After some feature engineering and joining, the following fields are returned.

Field	Data Type	Description
Title	character	Record ID
Created	date	Date record created
Accidental	character	Was the incident accidental
AimPoint	character	If a firearm was used, where was it aimed
Assaulted	character	Was the officer assaulted
AssessThreat	character	Description of threat assessment
BriefSummary	character	Brief summary of incident
CustodyNumber	character	Custody number
DateSubmitted	date	Date record was submitted
EventType	character	If the Use of Force was at an event, what was the event type
Firearm	character	If a firearm was used, was it aimed or fired
ForceUsedAgainst	character	How many people was force used against
ID	int	Record ID
ImpactBehaviours	character	What subject behaviours impacted whether Use of Force was used
IncidentDate	date	Date of the Use of Force incident
IncidentSetting	character	Setting in which incident took place
IncidentType	character	Type of incident where Use of Force was used
LevelOfInjury	character	If the subject sustained injury, were they minor or severe

Field	Data Type	Description
NPU	character	NPU of Use of Force incident
OperationalOrPublicOrderNumber	character	Operation or public order number
Outcome	character	The outcome of the Use of Force incident
ReasonForForce	character	Reasons Use of Force was used
SubdueSubject	character	Was the subject subdued
SubjectBehaviour	character	Subject behaviour
SubjectEthnicity	character	Ethnicity of the subject
SubjectGender	character	Gender of the subject
SubjectLevelOfInjury	character	Subject level of injury; none, minor, severe or death
TacticsUsed	character	Tactics used by officer
UseOfForceType	character	Type of incident; in custody, designated public order event or neither
WasForceUsed	character	Was force used
SubjectDOB	int	Subject date of birth
SubjectPerceivedAge	character	Subject percieved age group
OfficerCollarNumber	character	Officer collar number
ethnicity	character	Subject ethnic group
sex	character	Subject sex
officer_dob	date	Officer date of birth
date_joined	date	Date officer joined the force
officer_sex	character	Officer sex
officer_ethnic_origin	character	Officer ethnicity
last_date_paid	date	Date the officer was last paid, if they have left the force
last_date_worked	date	Date the officer last worked, if they have left the force
date_left	date	Date officer left the force
officer_age_end18	int	Officer age on 2018-12-31, in years
officer_experience_end18	int	Officer experience on 2018-12-31, in years
officer_age	int	Officer age on record created date

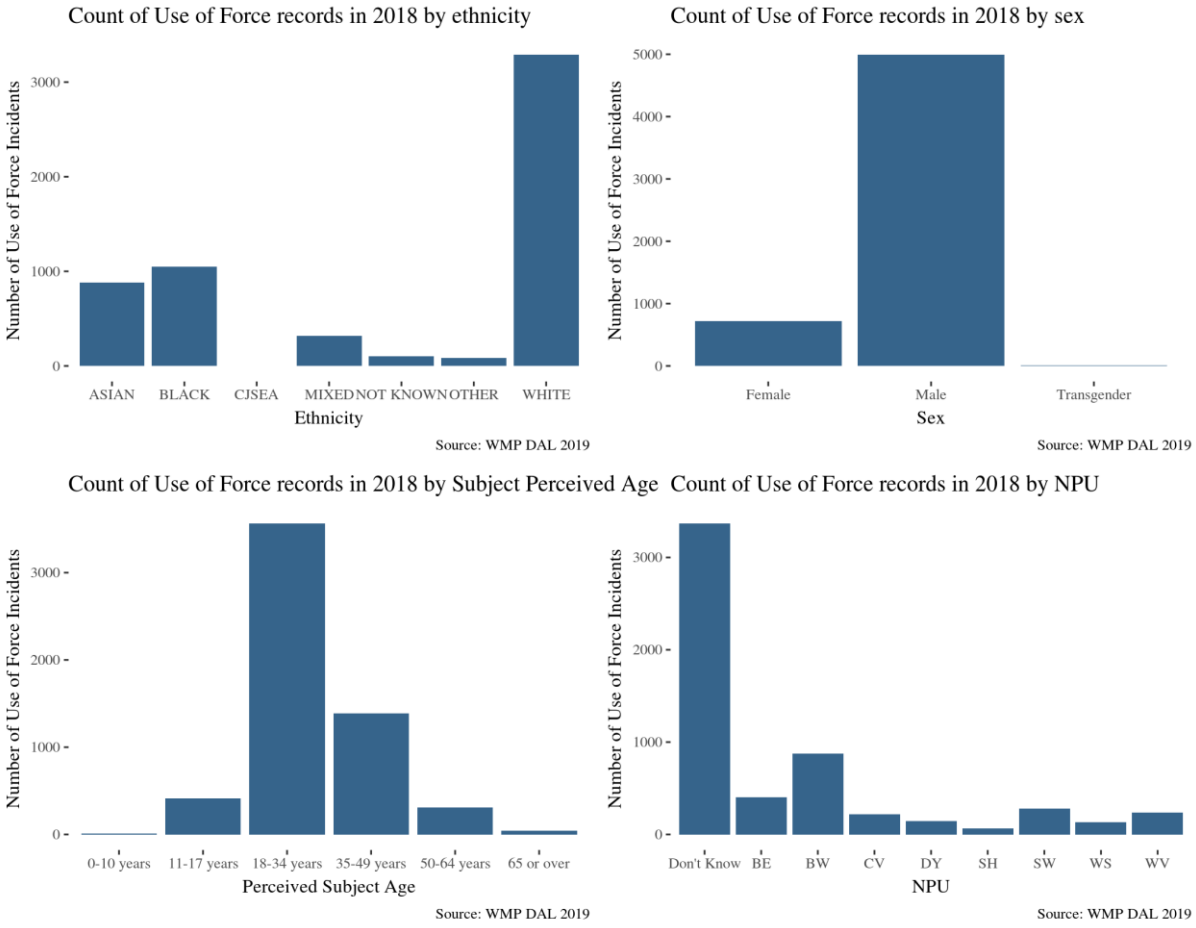
Field	Data Type	Description
officer_experience	int	Officer experience on record created date

Data dictionary of Use of Force data returned

7.2 Results

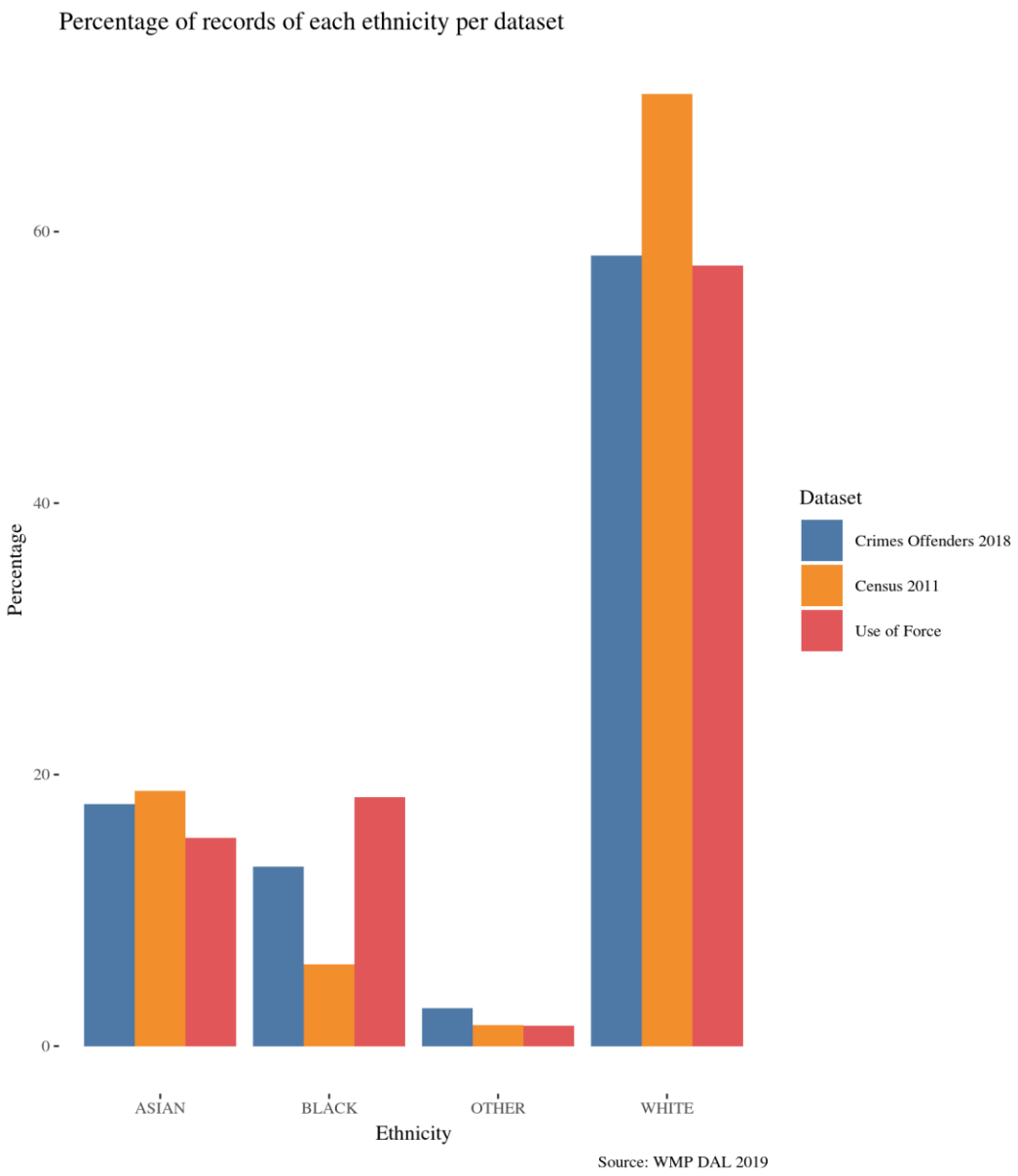
The Use of Force data cover a number of contexts including arrest, custody, keeping the peace, etc. In general therefore, population estimates are used for the reference populations.

The first plot is split into quadrants, the upper-left quadrant is the count of records of each subject ethnicity, the upper-right is the count of records of each subject sex, the lower-left is the count of records of each subject age group and the lower-right is the count of records in each NPU.



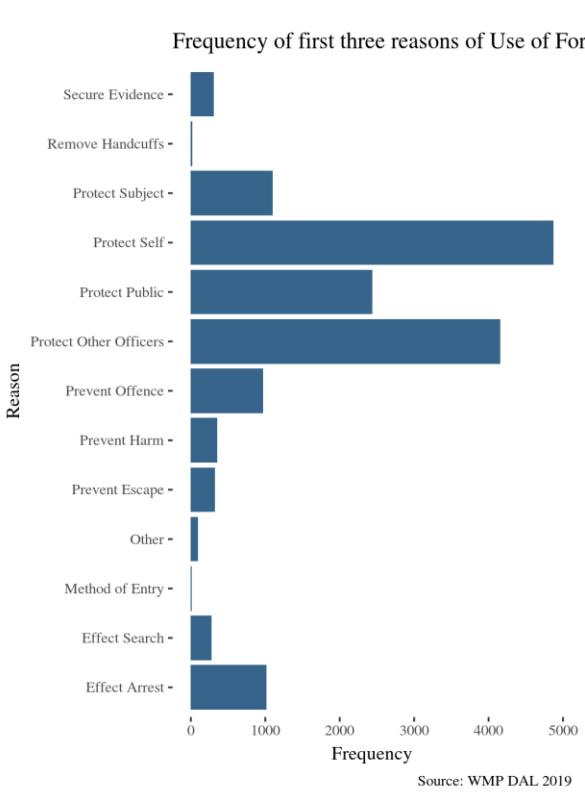
- There is a larger proportion of male subjects compared to the 2011 Census population, the percentage of male nominals in the 2011 Census is approximately 49%, this increases to 87% in the Use of Force data.
- The majority of Use of Force records do not have an associated NPU.

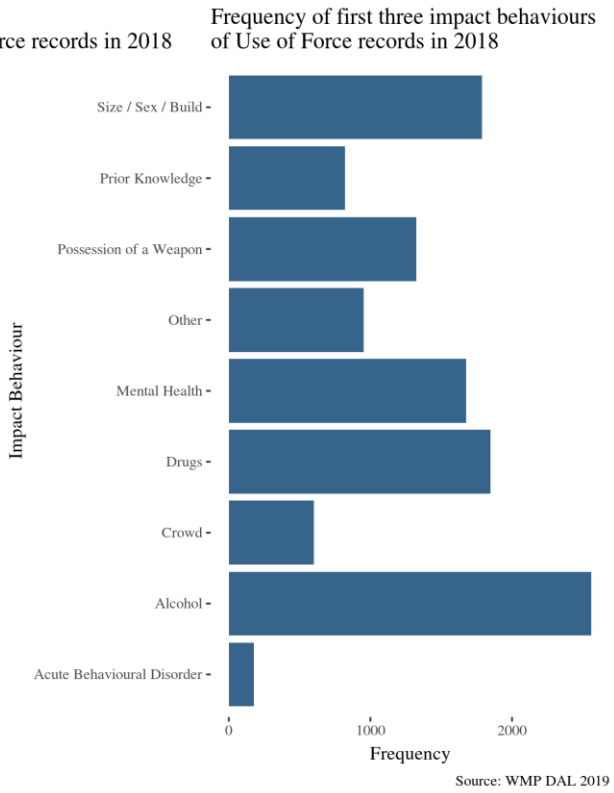
The below chart is a comparison of the percentage of nominals in the 2011 Census, Use of Force and the Crimes offenders data of each ethnicity.



From the above plot it is clear that the percentage of black nominals is higher in the Use of Force data (circa 18%) compared to Crimes offenders (circa 13%) and the 2011 Census (circa 6%).

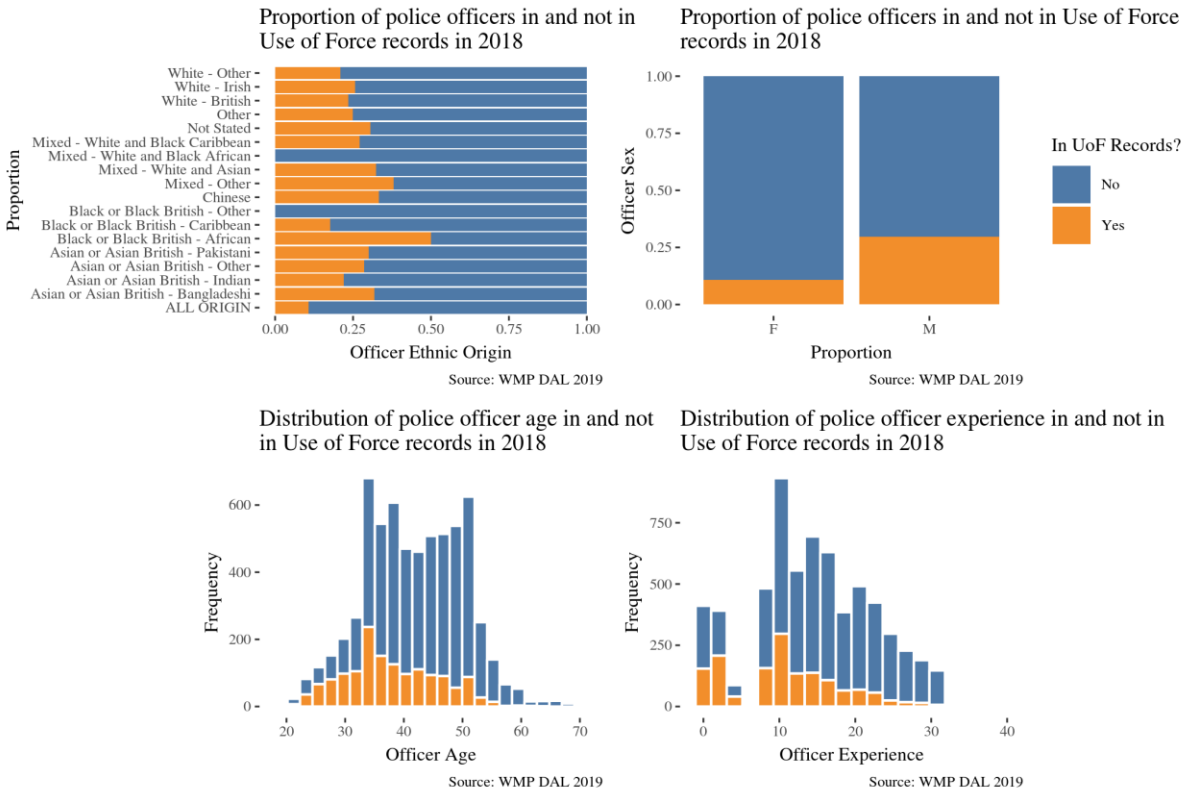
Within the Use of Force records, each individual record can (and often does) include multiple reasons for the use of force. Therefore, presence within the first three reasons listed within each record is used here. The second plot is split in two, the left side shows the number of records each reason has appeared in the first three reasons listed, the right side shows the number of records each impact behaviour has appeared in the first three impact behaviours (the impact behaviours are the factors which effect the decision to use force).





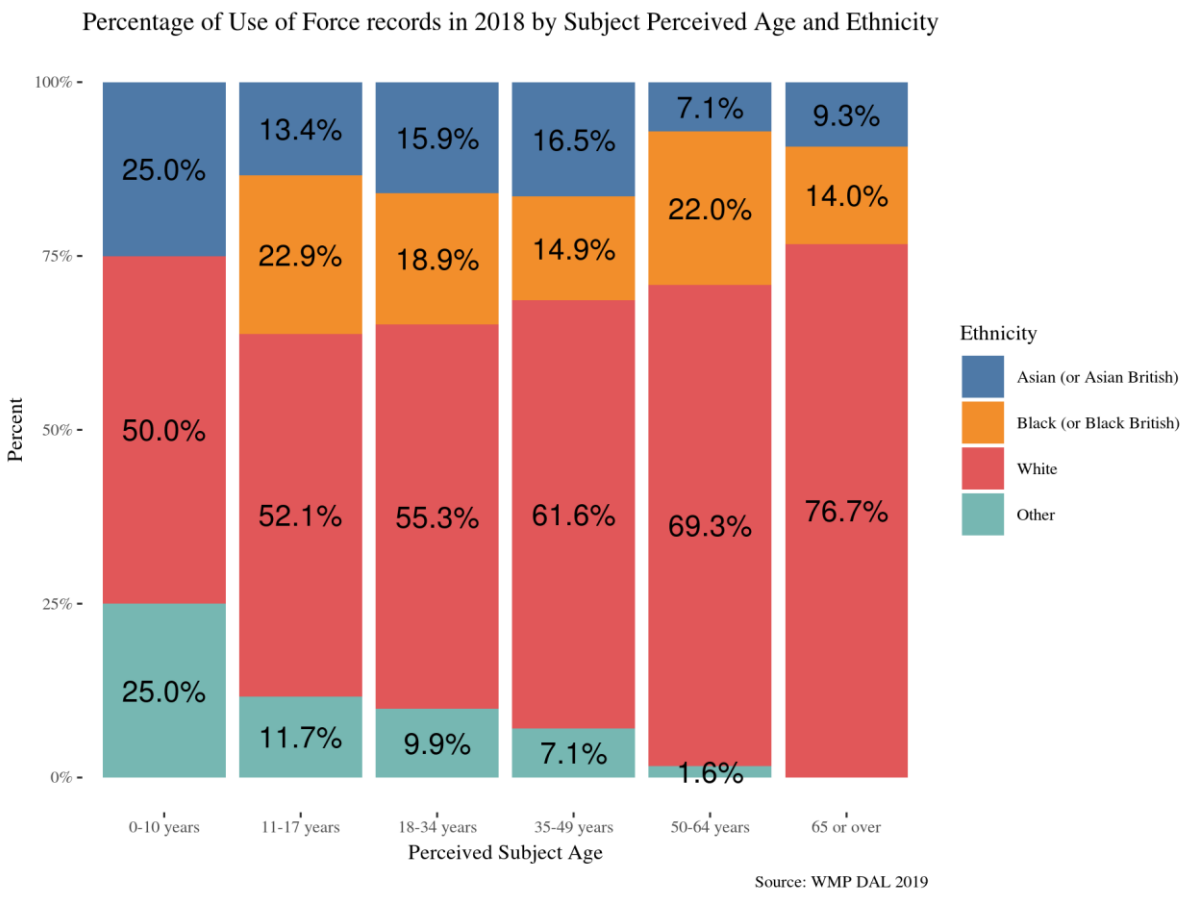
- The majority of the reasons are for protection - either of the officer, other officers, the public or the subject.
- The top three most common impact behaviours are alcohol, drugs and size / sex / build.

The following plot is split into quadrants, the upper-left shows the proportion of each officer ethnic origin in Use of Force records, the upper-right the proportion of officer sex in Use of Force records, the lower-left the distribution of officer age in and not in Use of Force records and the lower-right officer experience in and not in Use of Force records.



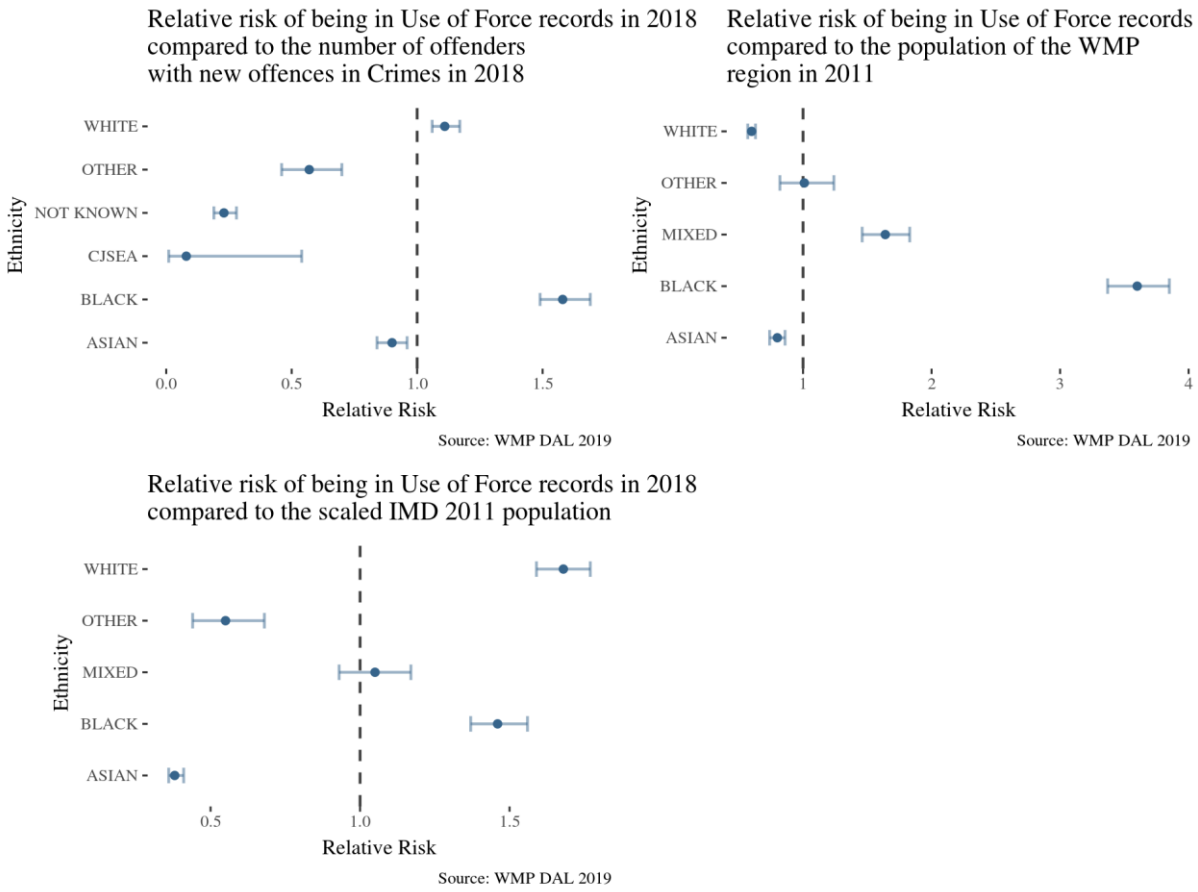
- There is a larger proportion of male officers in Use of Force records.
- The distribution of officer age in Use of Force records is skewed towards younger officers more than the officer population as a whole (younger officers are more likely to be in response and so affect an arrest than older officers).
- The distribution of officer experience is skewed towards less experienced officers more than the officer population (a similar process may be in play as noted in the point above).

The next plot shows the percentage of subjects of asian, black, white and other ethnic groups per age group.



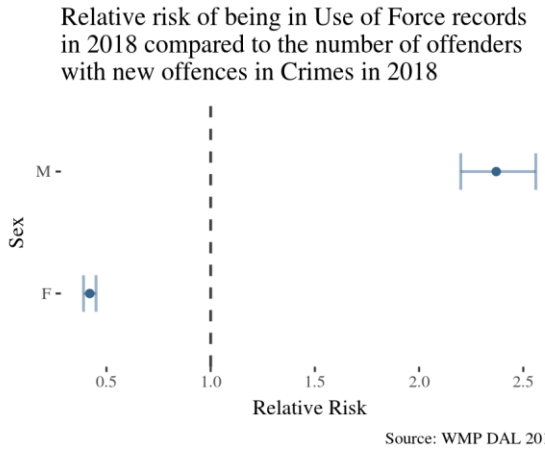
Note that there are only four subjects in the "0 - 10 years" old group. The percentage of white subjects increases as the subject age increases while the other three ethnicity groups generally decrease.

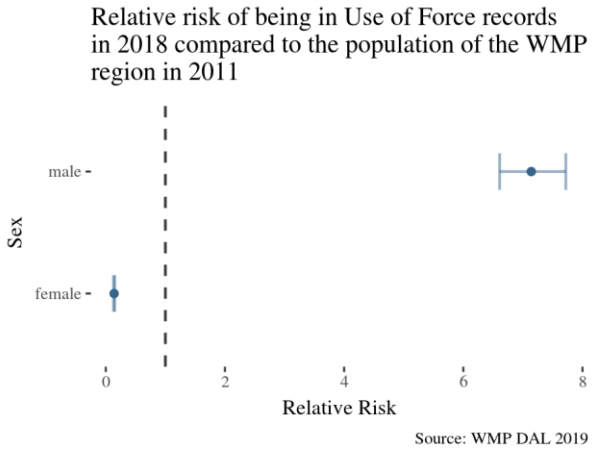
The below plot shows the relative risk in the Use of Force records for each ethnicity.

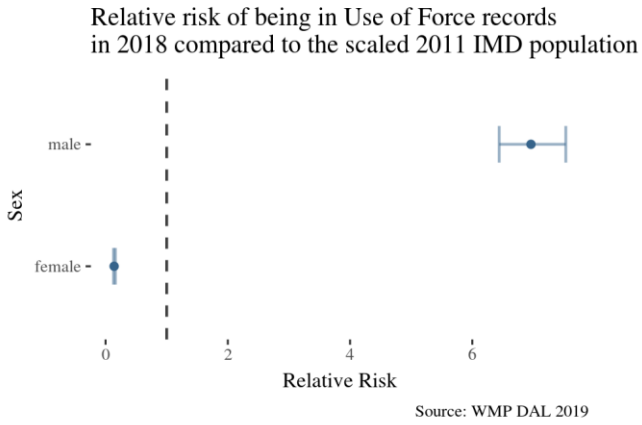


Black subjects generally have a larger relative risk of being in the Use of Force records across all three comparisons while asian subjects generally have a smaller relative risk. The relative risk for white subjects varies depending upon the population data used. As for the other systems this would seem to point that locational effects are at play.

The following plot shows the relative risk in the Use of Force records for each sex.

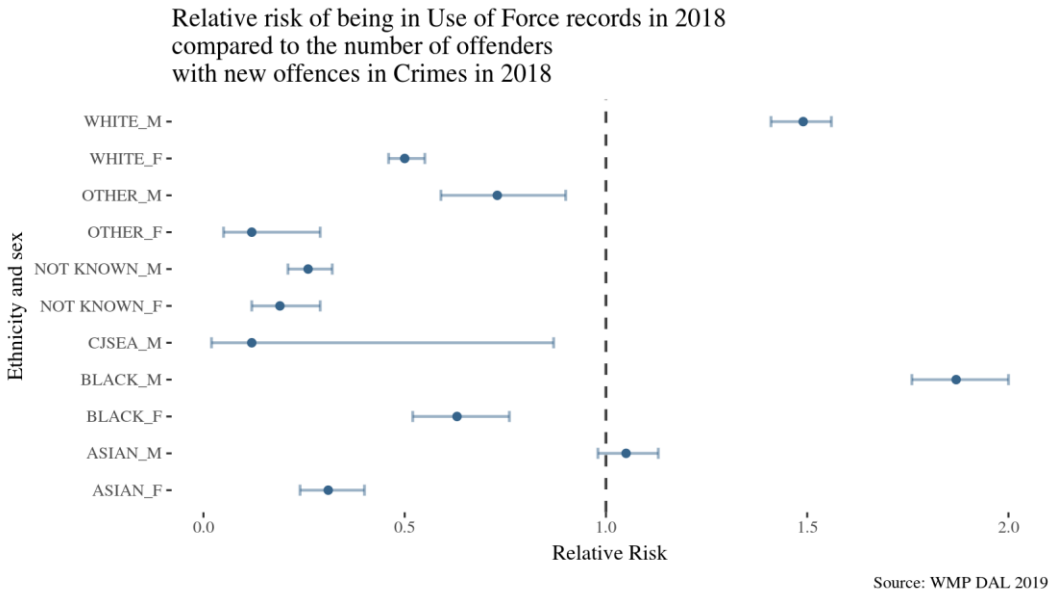






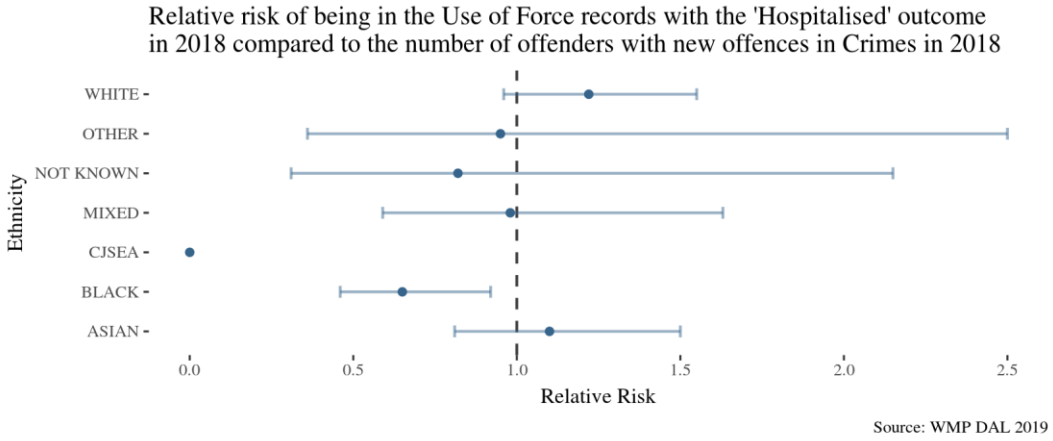
Male subjects have a larger relative risk of being in the Use of Force records compared to female subjects.

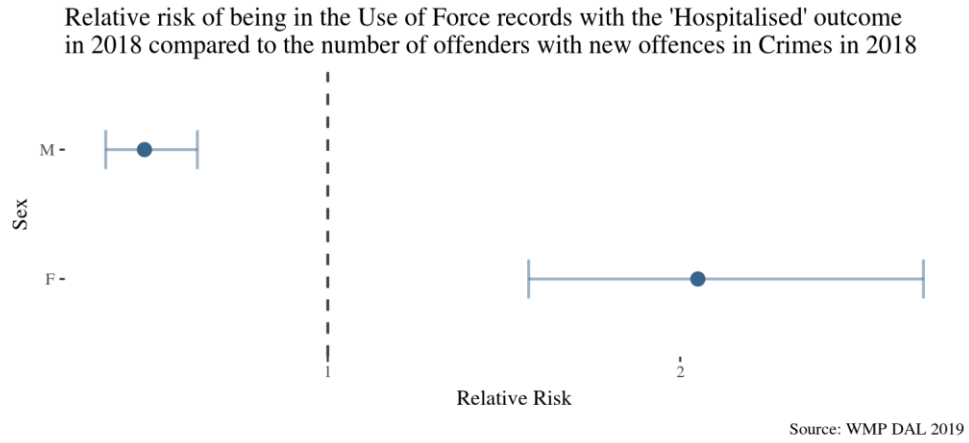
The below plot shows the relative risk in the Use of Force records for each ethnicity and sex combination.



Black and white male subjects have the larger relative risks while asian, white and black females have a smaller relative risk.

The final plot in this section is split in two, the top half shows the relative risk of the Use of Force record outcome being “Hospitalised” for each ethnicity, the bottom half shows the same for each sex.





Most of the relative risks for each ethnicity cross over one - meaning very little increase in relative risk. On the other hand, female subjects have a larger relative risk of being hospitalised as the outcome of the Use of Force compared to male subjects which is most likely due to existing conditions (including pregnancy) – i.e. they are generally taken to hospital not because of the use of force but because they needed to go to hospital prior to the interaction with Police.

8 TPO10

An extract of the TPO10 (traffic stop) data was uploaded for analysis.

8.1 Data

After some feature engineering the fields returned are as follows:

Field	Data Type	Description
id	character	Record ID
notice	character	Endorsable or non-endorsable notice
title	character	Nominal title
forename	character	Nominal forename
surname	character	Nominal surname
sex	character	Nominal sex
ethnicity	character	Nominal ethnicity
self_ethnicity	character	Ethnicity of officer
dob	date	Nominal date of birth
offence_date	date	Offence date
offence_code	character	Offence code
offence_desc	character	Offence description
speed	int	Speed of vehicle
street	character	Street where offence took place
locality	character	Locality where offence took place
town	character	Town where offence took place
vehicle_make	character	Make of vehicle in offence
vehicle_model	character	Model of vehicle in offence
vehicle_colour	character	Colour of vehicle in offence
officer	character	Officer collar number
force	character	Force recording offence
age	int	Age of nominal (years)
speeding	boolean	Whether the offence was speeding
speed_limit	character	If the offence was speeding, what was the speed limit
road_type	character	Road type; motorway, A road, other

Field	Data Type	Description
ethnic_group	character	Nominal ethnic group

Data dictionary of TPO10 data returned

The columns `speeding`, `speed_limit` and `road_type` were derived from `offence_desc` in the following way:

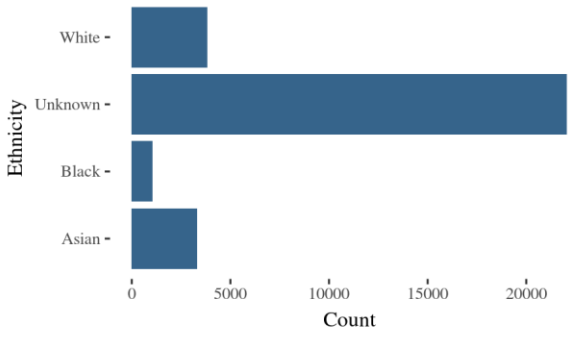
- `speeding` - this evaluates to TRUE if a string of characters of the form "[0-9][0-9] mph" was found in the `offence_desc` after it was converted to all lower case.
- `speed_limit` - if a string of the form "[0-9][0-9] mph" was found in the `offence_desc` after it was converted to all lower case, then the **value** of this column is the string found. If no string is found then the column value is "Not speeding".
- `road_type` - this column has three values depending on `offence_desc`:
    1. if a string of the form "m[0-9]" or "(m)" is found, the value is "M" (motorways),
    2. if a string of the form "a[0-9]" is found, the value is "A" (A roads),
    3. if none of the above strings are found, the value is "Other" (other roads not falling into motorways or A roads).

Note that personal data is not kept for records older than 2017, hence when examining details such as nominal age, sex and ethnicity the data will be filtered to records after 2016.

8.2 Results

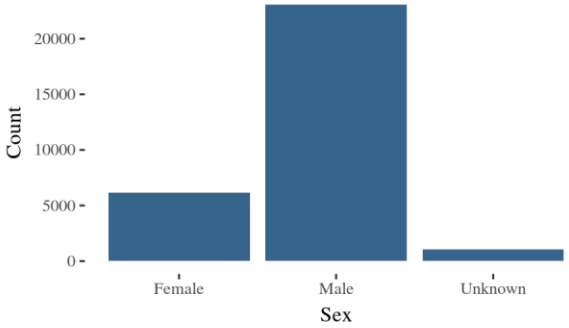
The first plot is split into quadrants, the upper-left shows the number of records for each ethnicity, the upper-right shows the number of records for each sex, the lower-left shows a histogram of age in the records, the lower-right shows a histogram of speeds for speeding offences.

Count by Ethnicity in traffic stop records from 2017 to 2019 inclusive



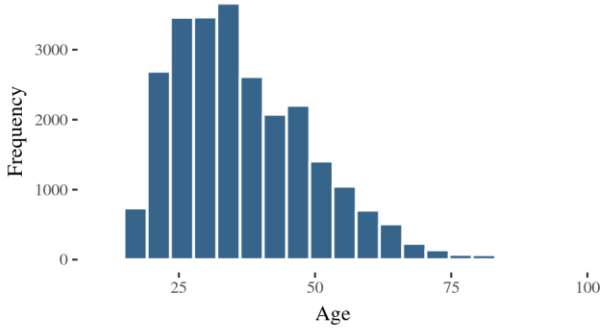
Source: WMP DAL 2019

Count by sex in traffic stop records from 2017 to 2019 inclusive



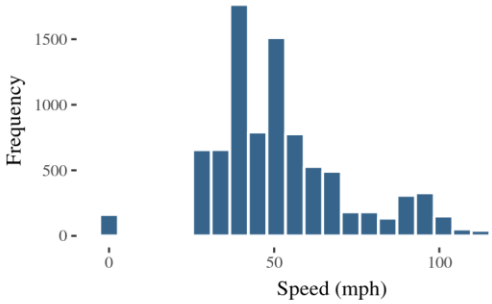
Source: WMP DAL 2019

Histogram of age in traffic stop records from 2017 to 2019 inclusive



Source: WMP DAL 2019

Histogram of speed in traffic stop records where the offence is speeding from 2013 to 2019 inclusive

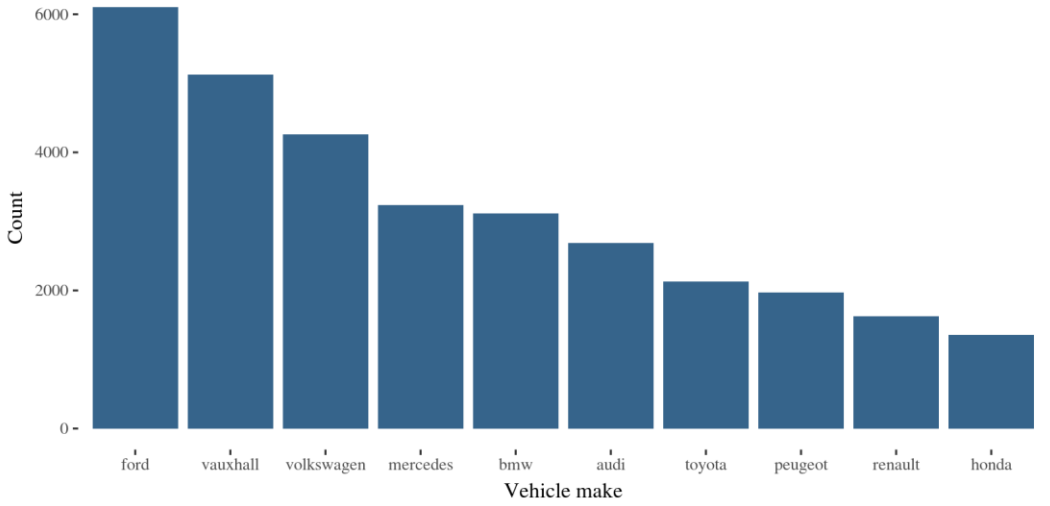


Source: WMP DAL 2019

- Most of the records have an unknown ethnicity (as personal attributes are deleted after 2 years, whilst the offence is kept for 6).
- There is a larger proportion of male nominals compared to the 2011 Census population.

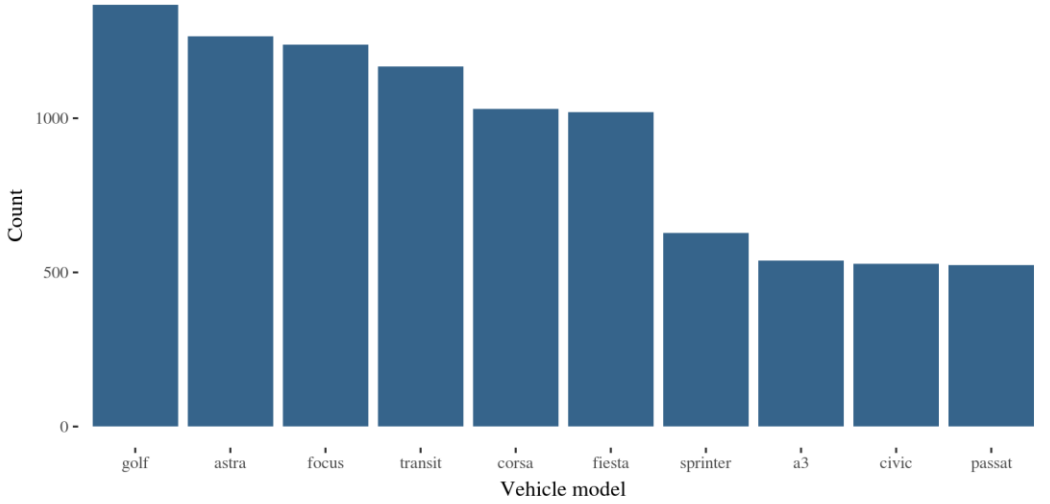
The second plot shows the count of records for each of the top ten vehicle make and models.

Count by vehicle make of traffic stop records from 2013 to 2019 inclusive



Source: WMP DAL 2019

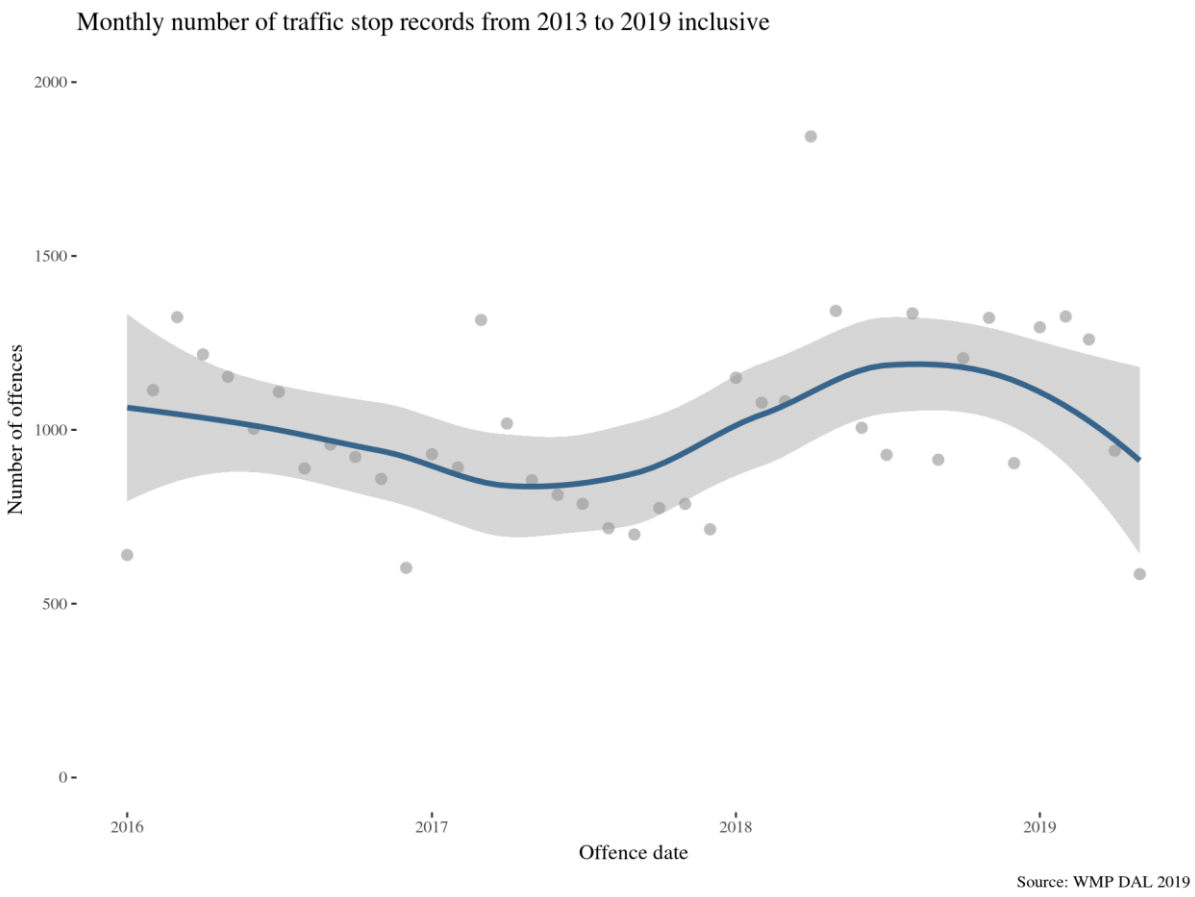
Count by vehicle model of traffic stop records from 2013 to 2019 inclusive



Source: WMP DAL 2019

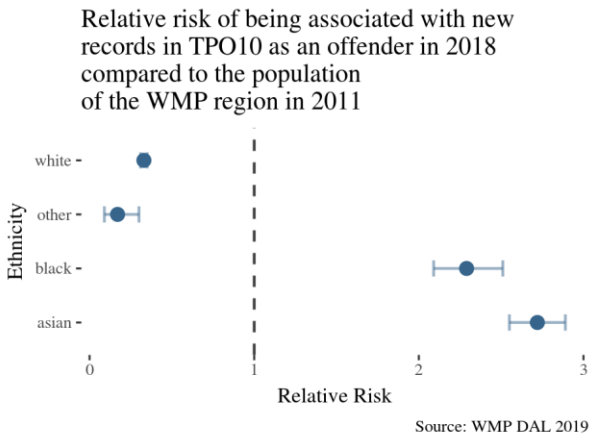
Ford is the most common make and Golf is the most common model.

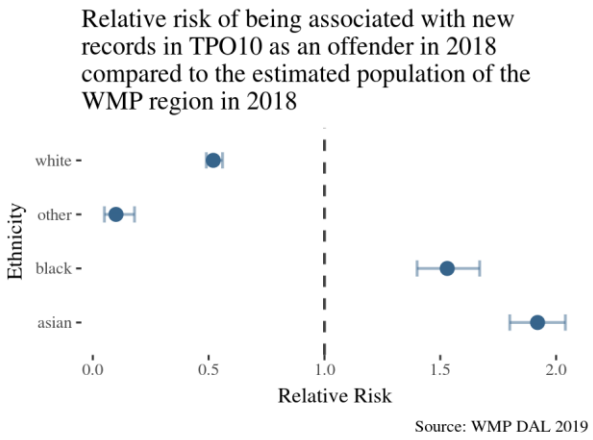
The following plot shows the total monthly number of traffic stops from 2013 to 2019.

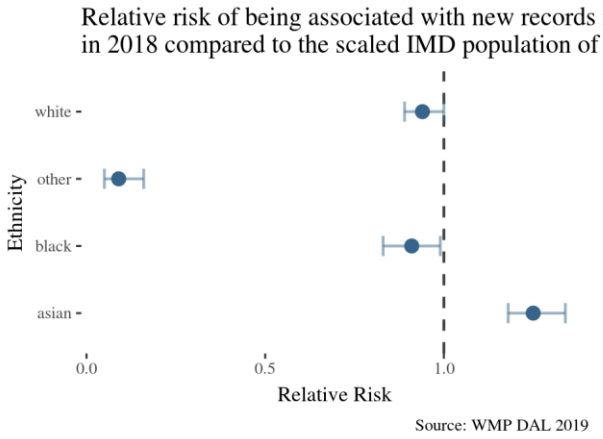


The monthly number of traffic stop records is relatively constant over time.

The final plot shows the estimate of relative risk in the traffic stop records per ethnicity.







Asian nominals have a larger relative risk of appearing in the traffic stop records across all comparisons whilst black nominals have a larger relative risk of appearing in the traffic stop records when the compared to the two population datasets. Below is a table summarising the number of records for each ethnicity.

Ethnicity	Number of TPO10 Records	Percentage of TPO10 Records
Asian	1,638	11.61%
Black	544	3.86%
CJSEA	15	0.11%
Other	11	0.08%
Unknown	9,969	70.65%
White	1,934	13.71%

Counts and percentages for each ethnicity in the traffic stop records where the offence was committed in 2018

9 Conclusion

To conclude, the main points from the report are:

- male nominals have a larger relative risk of being associated with new records in Crimes as offenders
- black nominals have a larger relative risk of entering the custody records from 2011 to 2018 compared to other ethnicities
- the number of Stop and Searches in an area is not only dependent on the population of those areas, but also on the activities occurring within those areas. This difference can lead to skews when calculating risk ratios (e.g. the LSOA covering the airport)
- black nominals have a larger relative risk of being in the Stop and Search records compared to other ethnicities
- male nominals have a larger relative risk of being in the Stop and Search and Use of Force records compared to female nominals
- asian nominals have a larger relative risk of being in the traffic stop records.

The analyses undertaken here should be taken as preliminary with a view to identifying further areas of research.

10 Appendix

10.1 Relative Risk Methodology

Suppose X is an event for which an estimate of relative risk is desired across a group G. The complement of X and G are denoted by X' and G' respectively. In words, X' is equivalent to the event X did not happen and G' is equivalent to the group is not G. The relative risk can be estimated by first constructing a contingency table:

	X	X'	Total
G	a	b	a + b
G'	c	d	c + d

An estimate of relative risk can then be calculated by

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)}$$

The confidence interval at significance level α is then calculated by

$$\ln(\widehat{RR}) \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}$$

11 Bibliography

gov.uk. 2011. “Police Powers to Stop and Search: Your Rights.” viewed 2019-08-12 <https://www.gov.uk/police-powers-to-stop-and-search-your-rights>.

National Statistics, Office for. 2011. “English Indices of Deprivation 2010.” viewed 2019-07-15 <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>.

———. 2012. “2011 Census: Population and Household Estimates for Small Areas in England and Wales, March 2011.” viewed 2019-07-15 <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23>.

Office, Home. 2018. “Police Use of Force Statistics, England and Wales: April 2017 to March 2018.” viewed 2019-08-13 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/764894/police-use-of-force-apr2017-mar2018-hosb3018.pdf>.