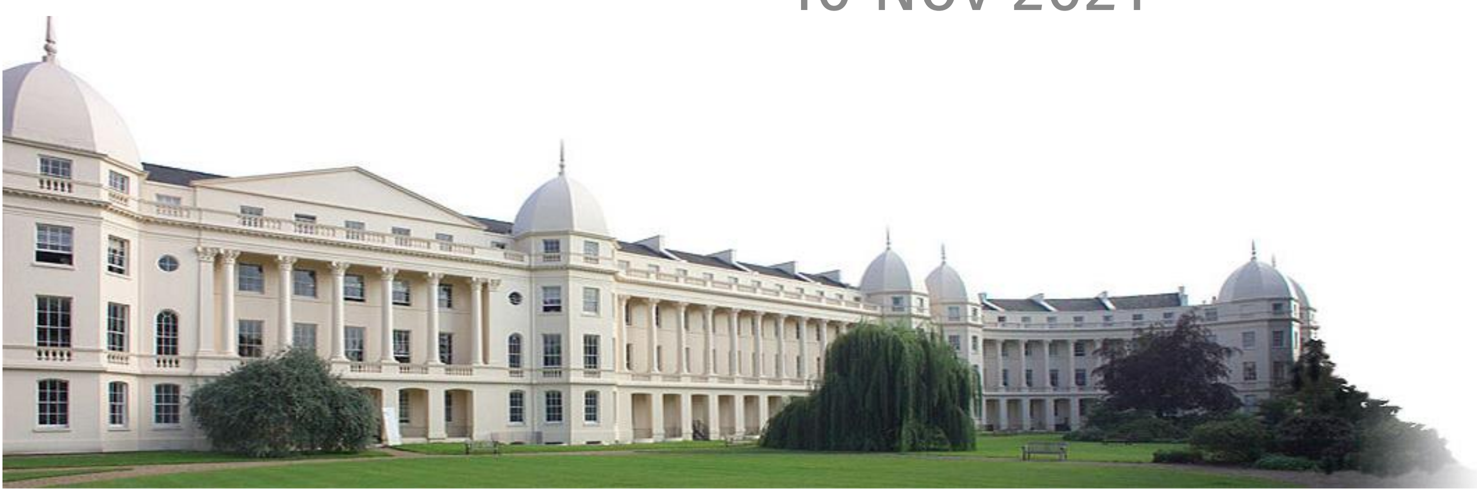
London Business School

Data Visualisation AM10
Session 1 Visualisations

Kostis Christodoulou

10 Nov 2021



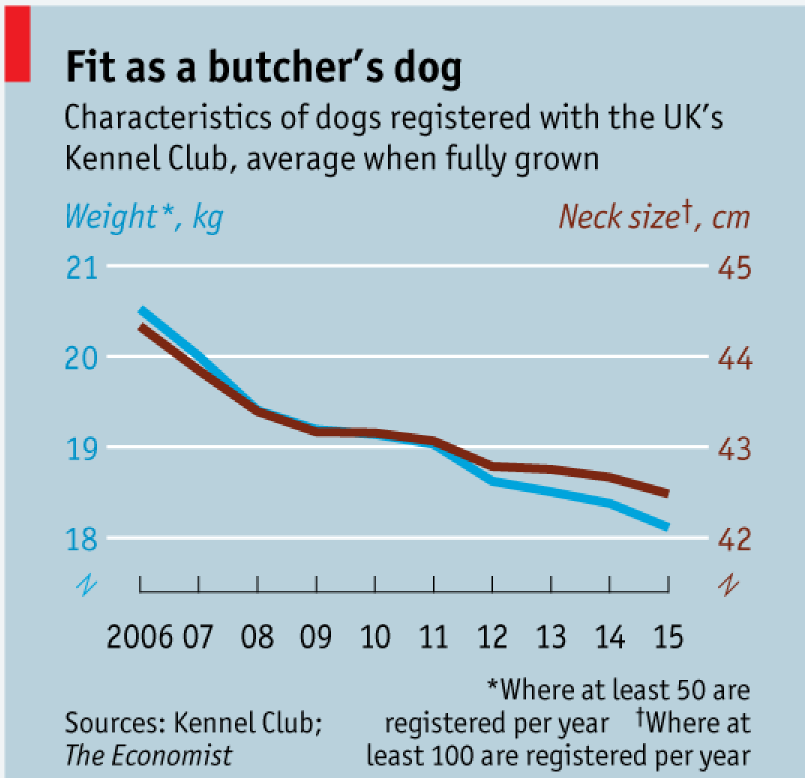
London Business School

Why not use double y-axes?

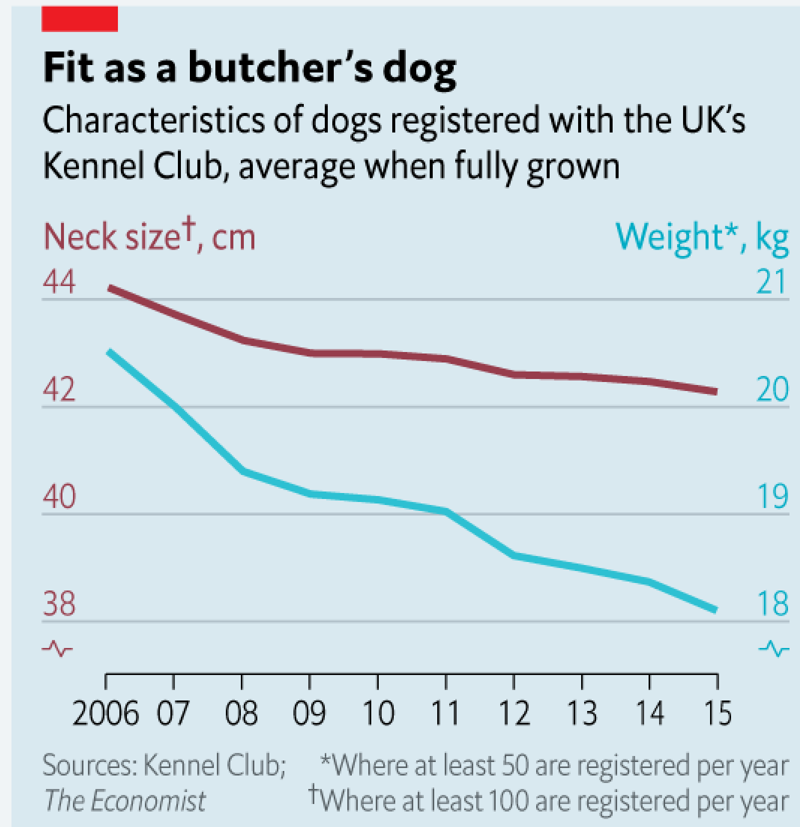
00_secondary_axis.R

If you choose where the y-axes start and stop, you can make the two trends to line up however you want!

Original



Better

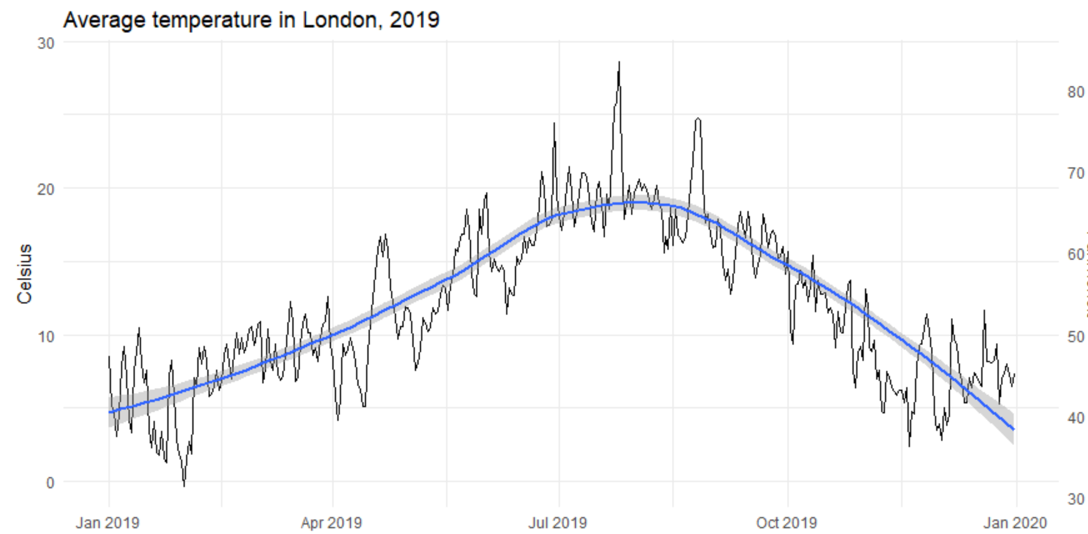


https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368

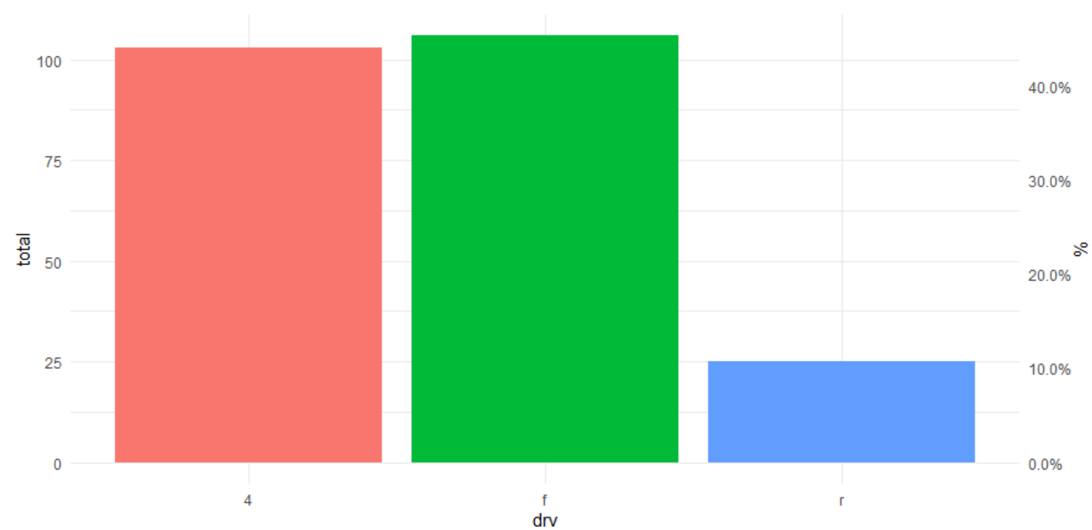
OK when both axes measure the same thing

00_secondary_axis.R

```r
# from london_bikes data, that contains avg_temp
bikes %>%
  filter(year == 2019 ) %>%
  ggplot(aes(x = date, y = avg_temp)) +
  geom_line() +
  geom_smooth() +
  labs(
    title = "Average temperature in London, 2019",
    x = NULL,
    y= "Celsius")+
  scale_y_continuous(
    sec.axis =
      sec_axis(trans = ~ (1.8 * .) +32,
               name = "Fahrneheit")
  ) +
  theme_minimal()+
  NULL
```




```r
car_counts <- mpg %>%
  group_by(drv) %>%
  summarize(total = n())

total_cars <- sum(car_counts$total)

ggplot(car_counts,
       aes(x = drv, y = total,
           fill = drv)) +
  geom_col() +
  scale_y_continuous(
    sec.axis = sec_axis(
      trans = ~ . / total_cars,
      labels = scales::percent,
      name = "%")
  ) +
  guides(fill = FALSE)+
  theme_minimal()+
  NULL
```



Assignment 1 thought piece

AM10 Reflection Paper

Why do we visualise data? What makes a great visualisation?

We visualise data because the human brain is more effective at detecting patterns, trends and outliers and making clear inferences when presented with information visually, than in almost any other way. A

Pareidolia: the tendency for incorrect perception of a stimulus as an object, pattern or meaning known to the observer, such as seeing shapes in clouds, seeing faces in inanimate objects or abstract patterns, or hearing hidden messages in music.





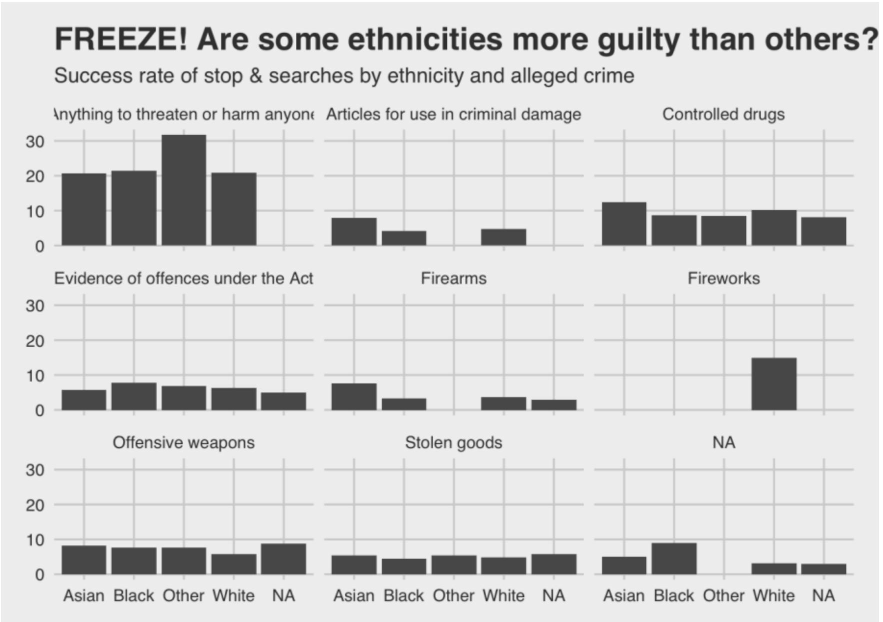
Assignment 1 visualisations

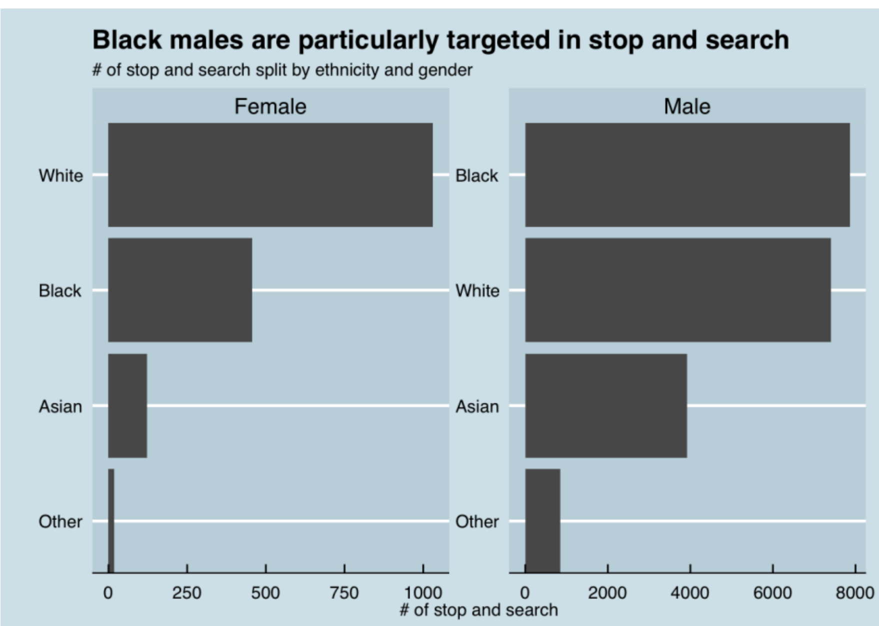
```r
stop_and_search <- stop_and_search %>%

  # Renaming a few variables which included spaces
  rename(PartOfPolicingOperation = `Part of a policing operation`,
         PolicingOperation = `Policing operation`,
         AgeRange = `Age range`,
         EthnicitySelfDefined = `Self-defined ethnicity`,
         EthnicityOfficerDefined = `Officer-defined ethnicity`,
         ObjectOfSearch = `Object of search`,
         OutcomeLinkedToObjectOfSearch = `Outcome linked to object of search`,
         RemovalOfMoreThanJustOuterClothing = `Removal of more than just outer clothing`) %>%

  mutate(

    # Create a new variable for the time of day during which the stop and search took place
    TimeOfDay = case_when(hour(Date) >= 23 | hour(Date) < 5 ~ "Night",
                          hour(Date) >= 17 ~ "Evening",
                          hour(Date) >= 11 ~ "Day",
                          hour(Date) >= 5 ~ "Morning"),

    # Create a new variable that only stores the date without time
    Day = date(Date)
  ) %>%
```

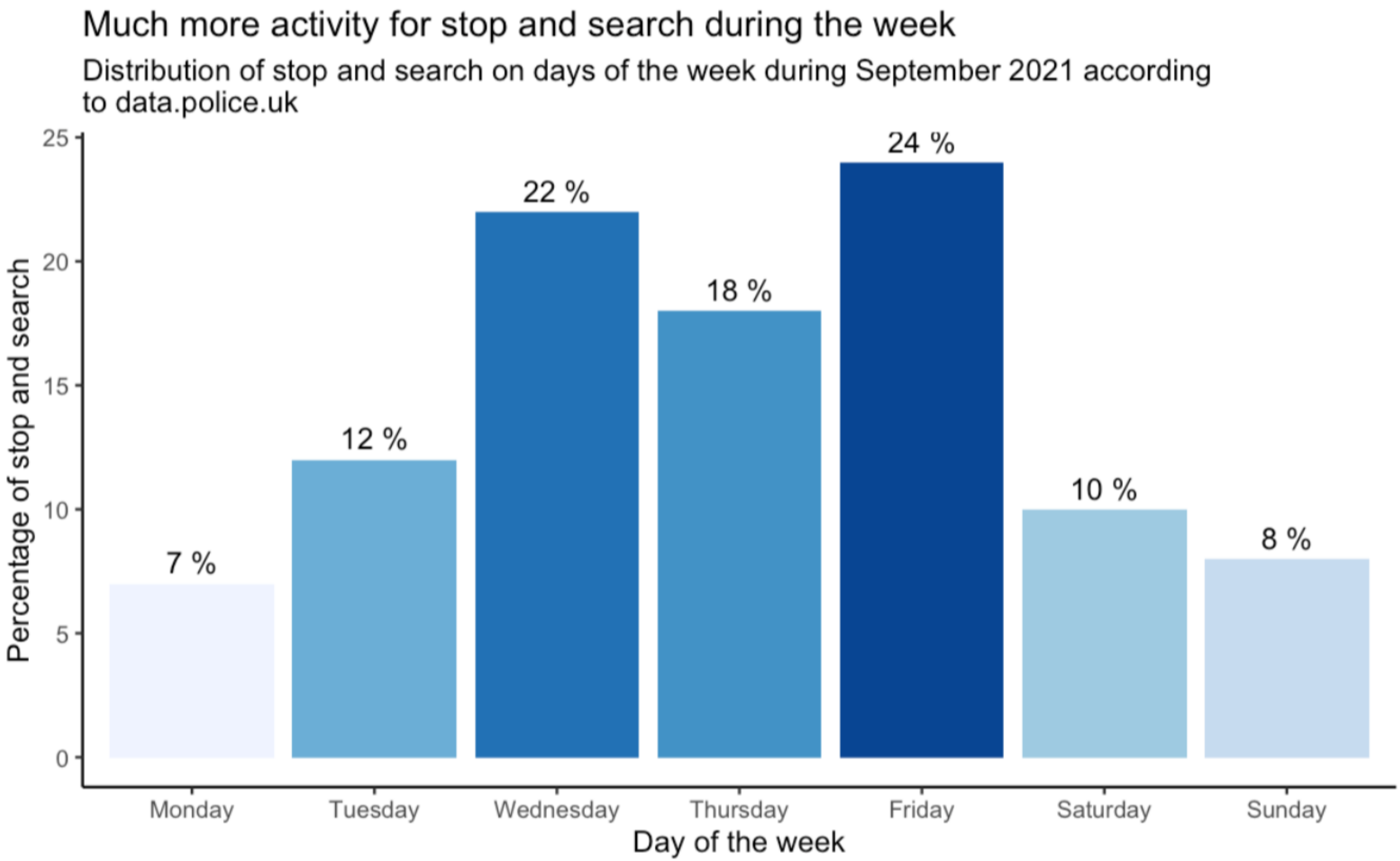
janitor::clean_names()

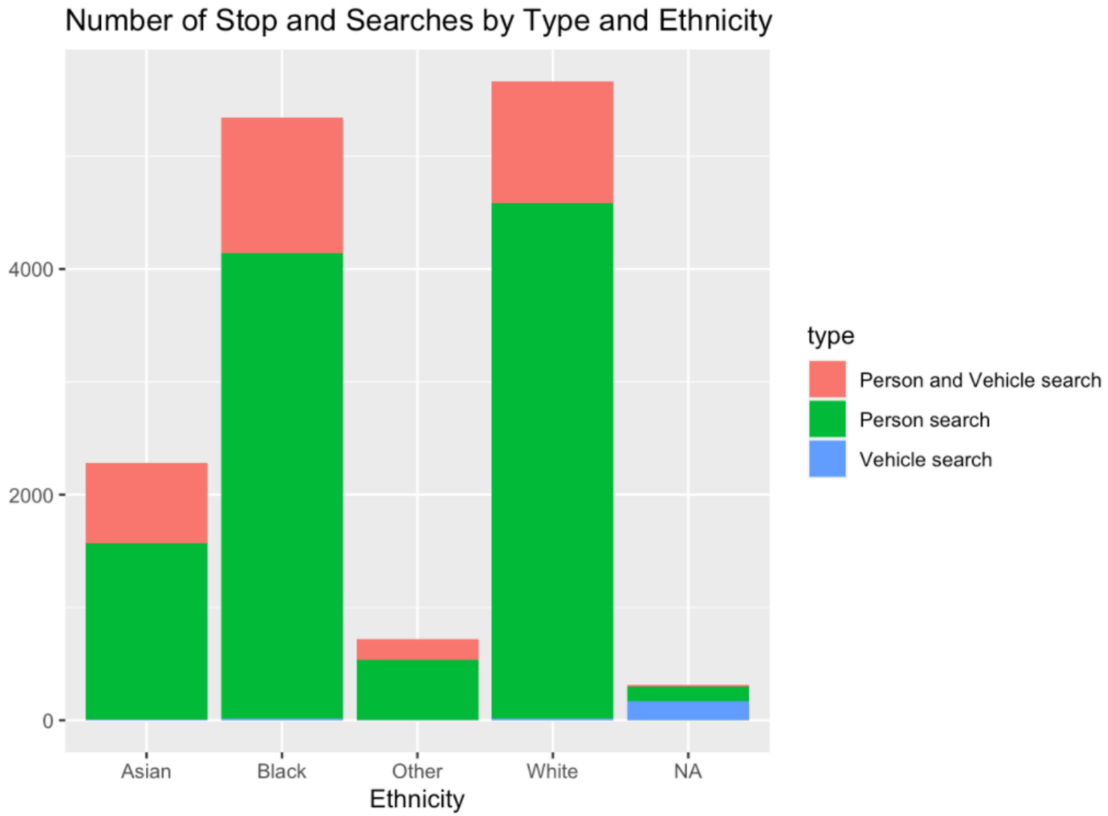
lubridate

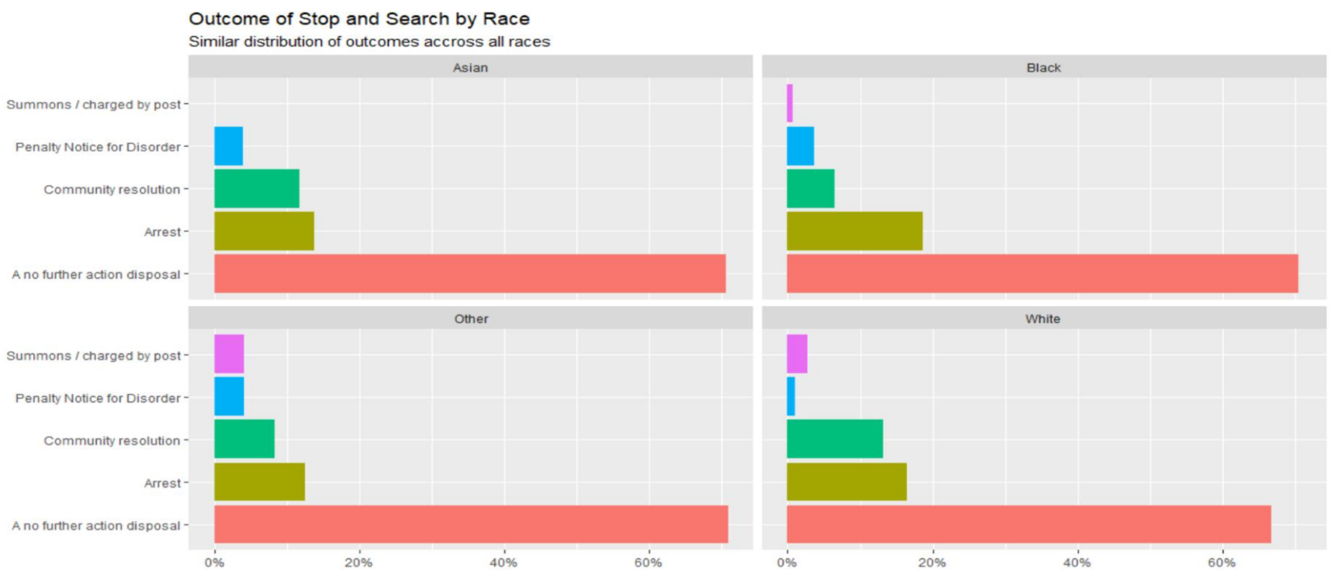
Assignment 1 visualisations

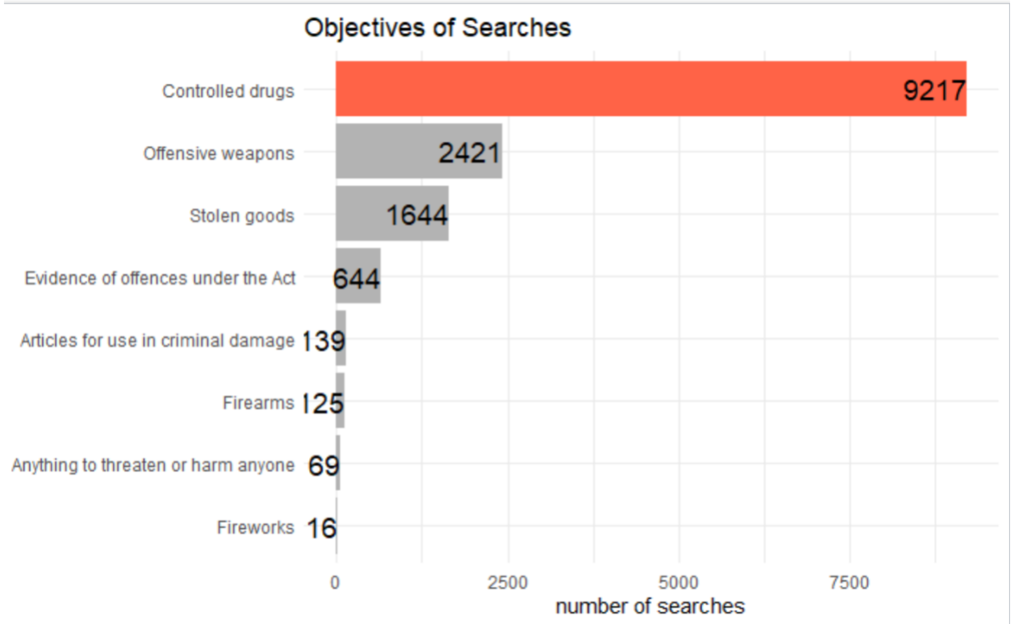


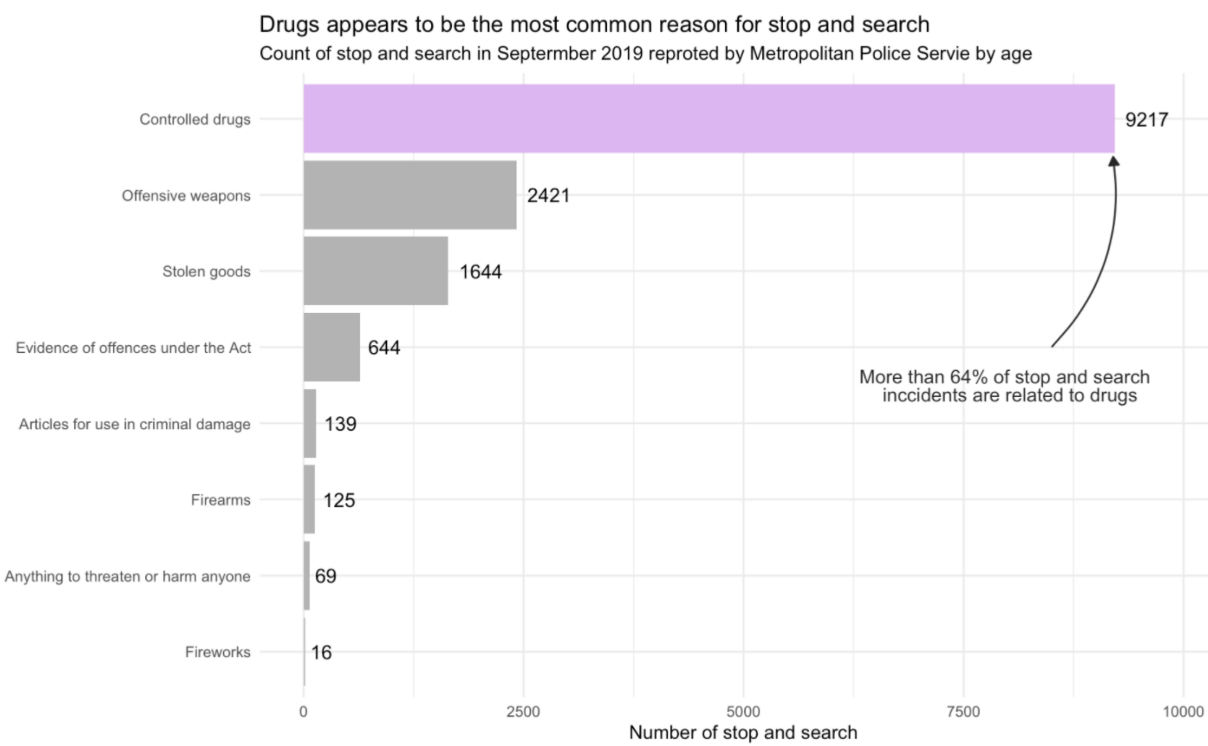



Much more activity for stop and search during the week
Distribution of stop and search on days of the week during September 2021 according to data.police.uk
Percentage of stop and search
25
20
15
10
5
0
7 %
12 %
22 %
18 %
24 %
10 %
8 %
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the week


Number of Stop and Searches by Type and Ethnicity
4000
2000
0
Asian
Black
Other
White
NA
Ethnicity
type
Person and Vehicle search
Person search
Vehicle search

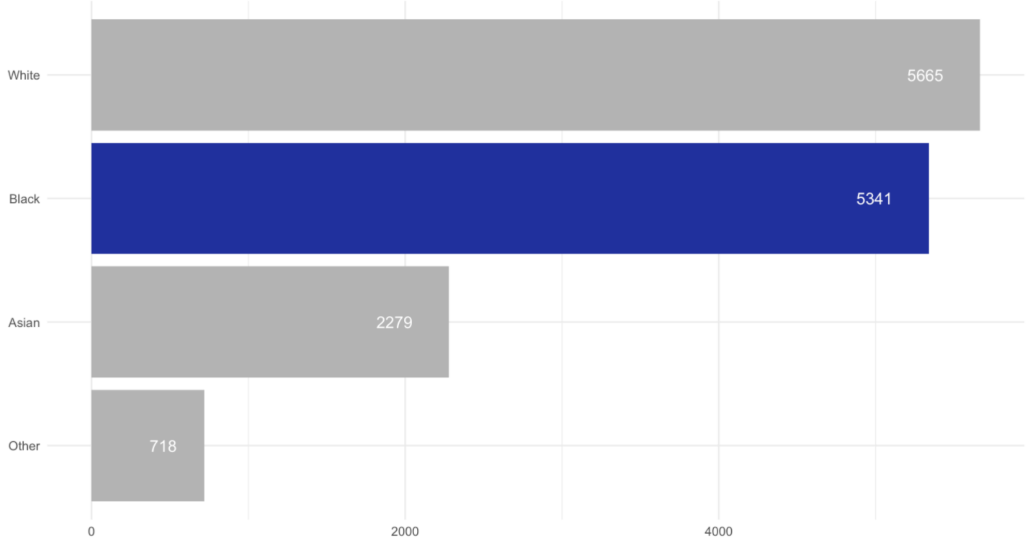

Outcome of Stop and Search by Race
Similar distribution of outcomes accross all races
Asian
Black
Other
White
Summons / charged by post
Penalty Notice for Disorder
Community resolution
Arrest
A no further action disposal
0%
20%
40%
60%


Objectives of Searches
Controlled drugs
Offensive weapons
Stolen goods
Evidence of offences under the Act
Articles for use in criminal damage
Firearms
Anything to threaten or harm anyone
Fireworks
9217
2421
1644
644
139
125
69
16
0
2500
5000
7500
number of searches

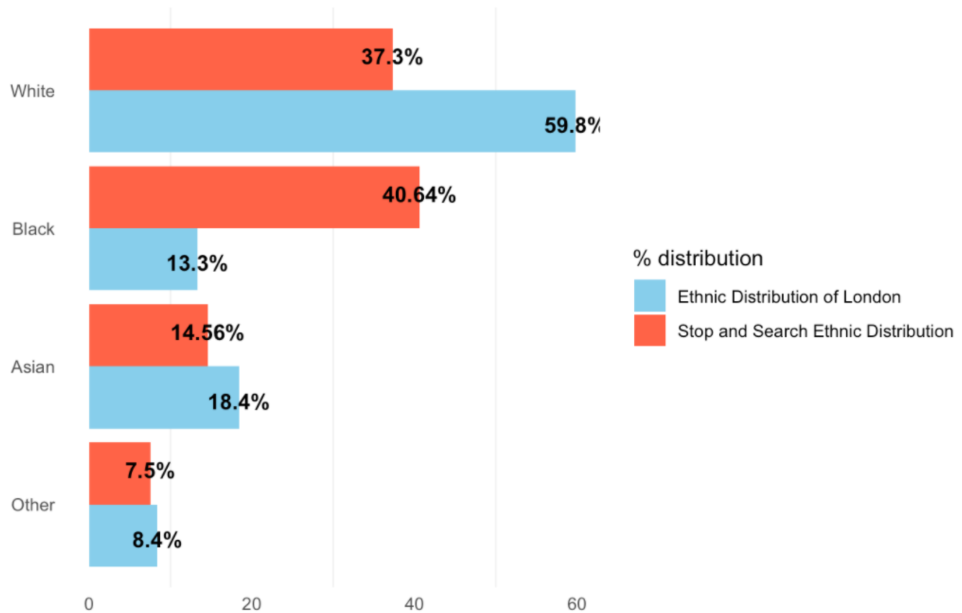

Drugs appears to be the most common reason for stop and search
Count of stop and search in Septermber 2019 reproted by Metropolitan Police Servie by age
Controlled drugs
Offensive weapons
Stolen goods
Evidence of offences under the Act
Articles for use in criminal damage
Firearms
Anything to threaten or harm anyone
Fireworks
9217
2421
1644
644
139
125
69
16
More than 64% of stop and search
inccidents are related to drugs
0
2500
5000
7500
10000
Number of stop and search

Black people appear to be overfrequently stopped when compared to population proportion

Count of stop and search in Septermber 2019 reproted by Metropolitan Police Servie by race



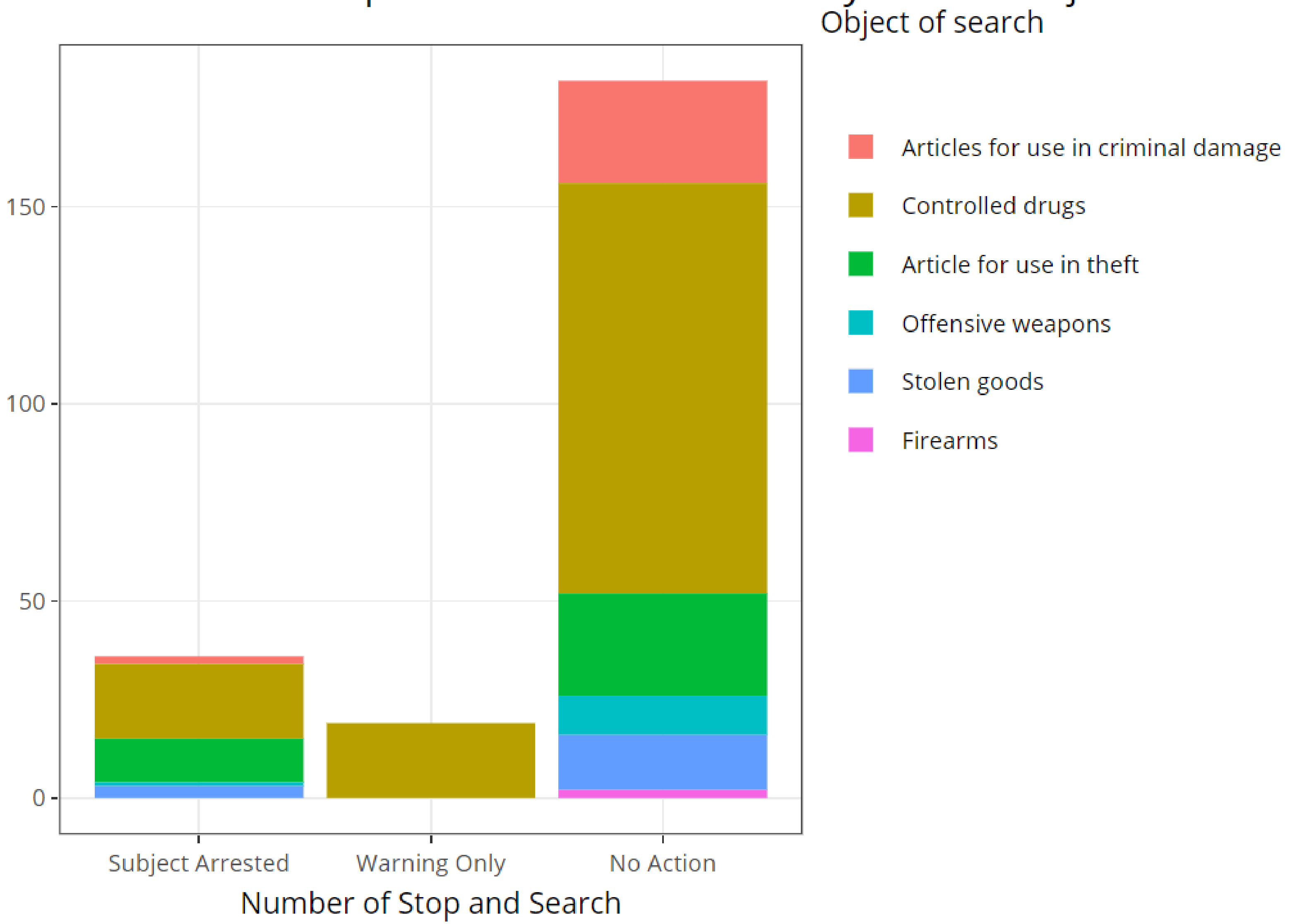
40% of Stop and Searches conducted on 13% of Londons population



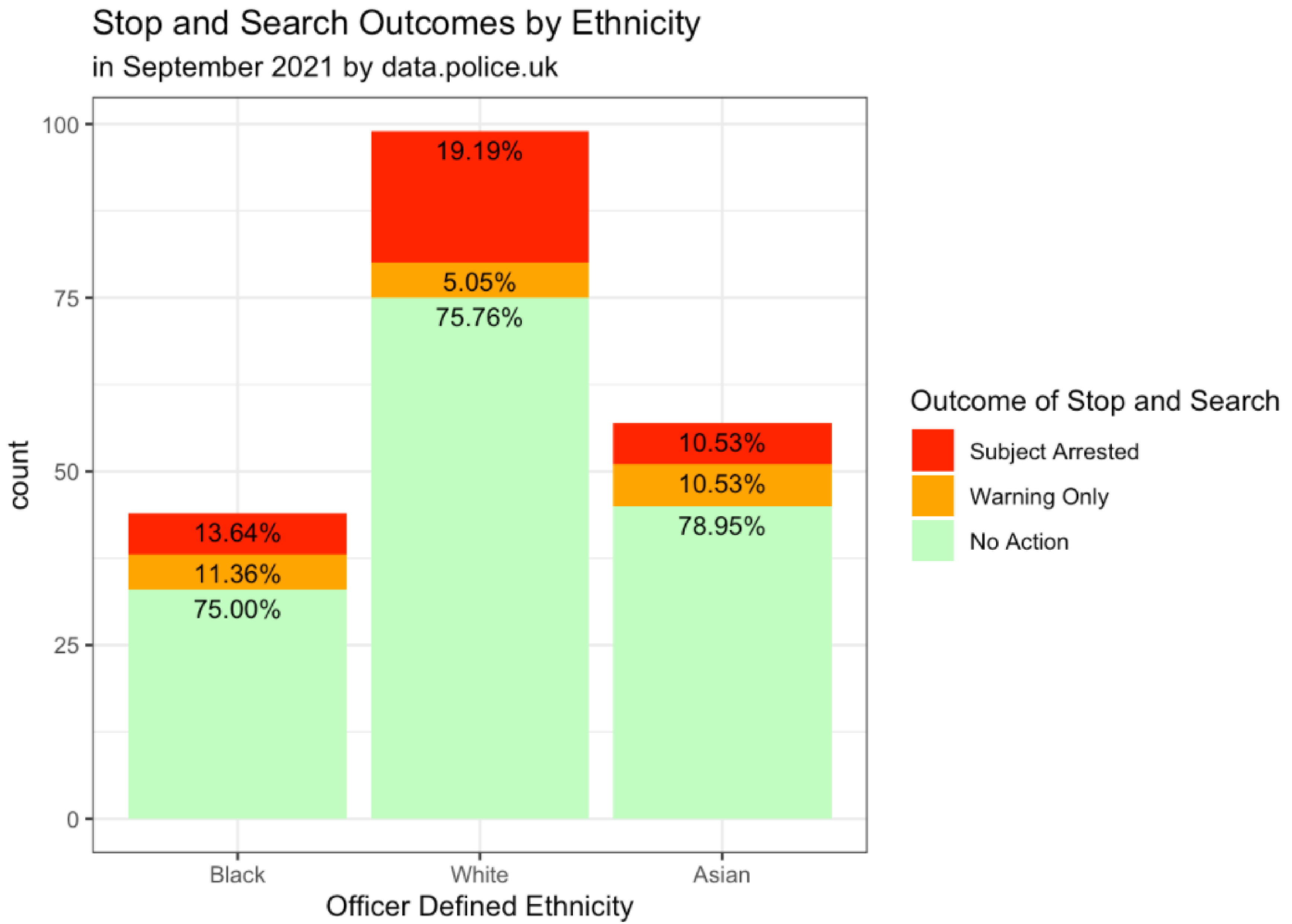
NOTE: Ethnicity Breakdown of London from Wikipedia

London Business School

Outcome of Stop and Search in London by Search Object

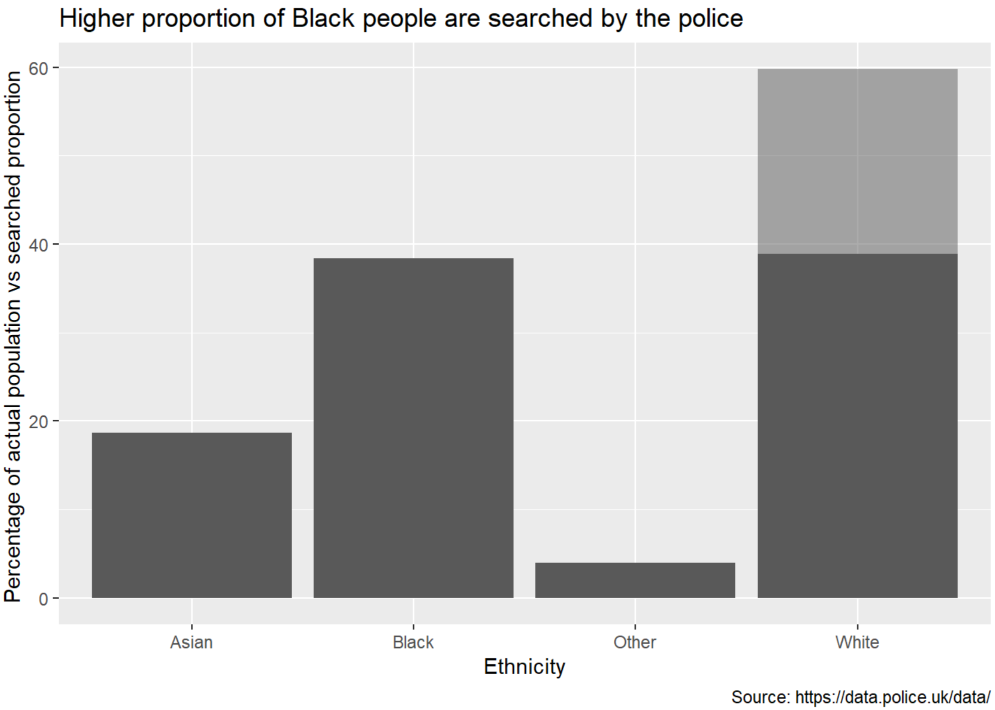


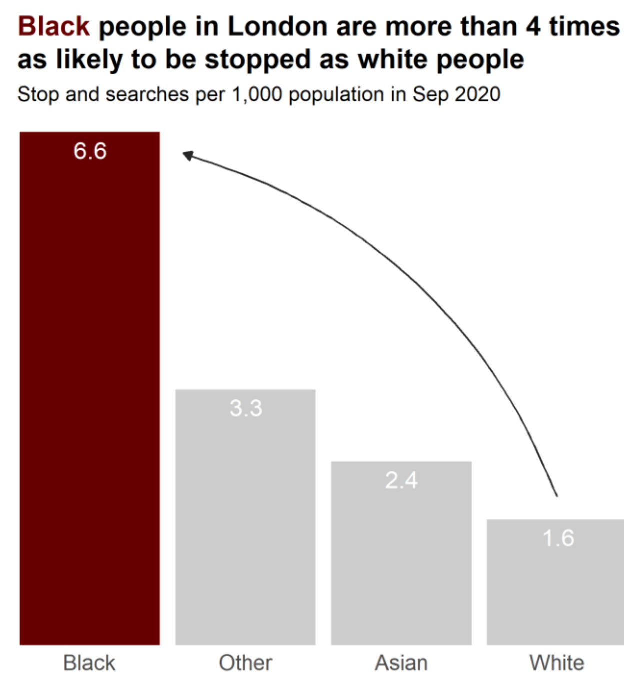
London Business School

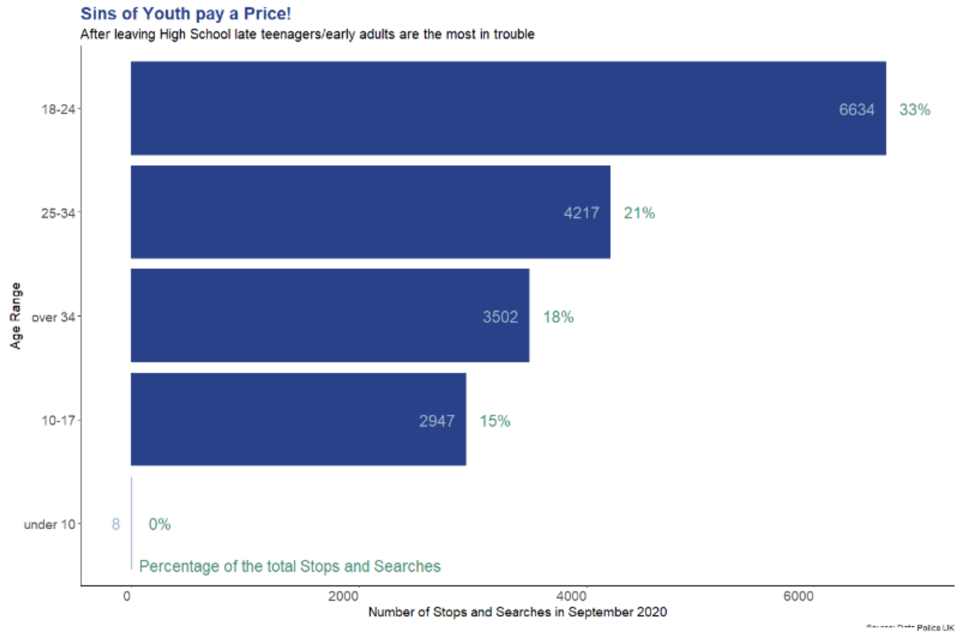


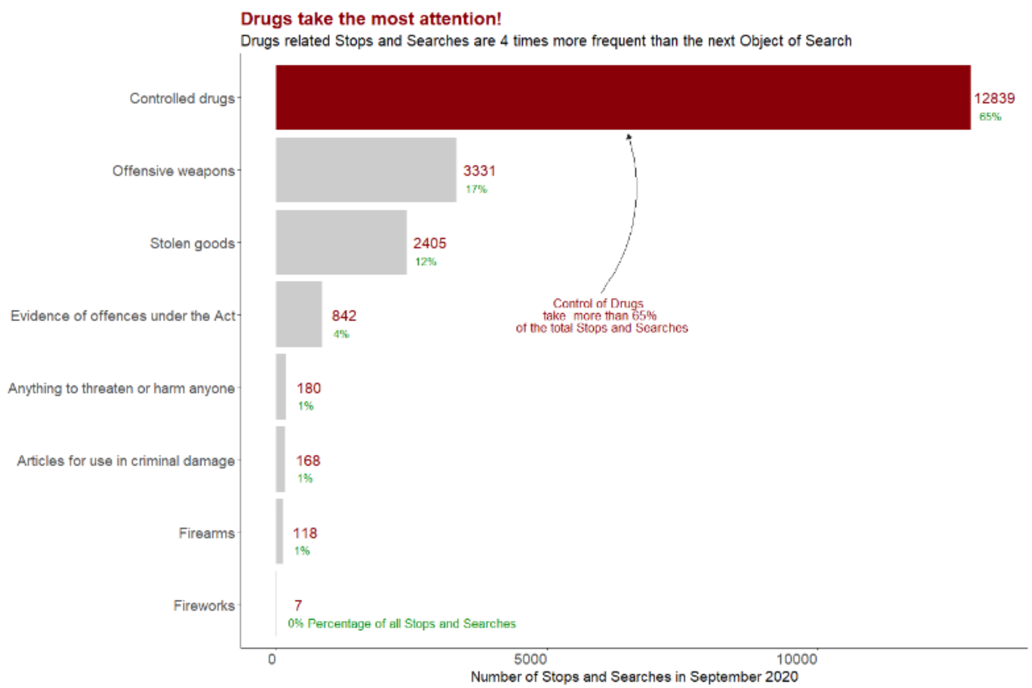
```
##   ethnicity searchpercent actualper
## 1     Asian      18.69344      18.4
## 2     Black      38.39335      13.3
## 3     Other       4.00277       3.4
## 4     White      38.91043      59.8
```

```
plot4 <- ggplot(percomp, aes(ethnicity, actualper)) + geom_col(alpha=0.5)+ geom_col(aes(ethnicity, searchper))+
  labs(title = "Higher proportion of Black people are searched by the police",
       caption = "Source: https://data.police.uk/data/",
       x = "Ethnicity", y= "Percentage of actual population vs searched proportion ")
plot4
```

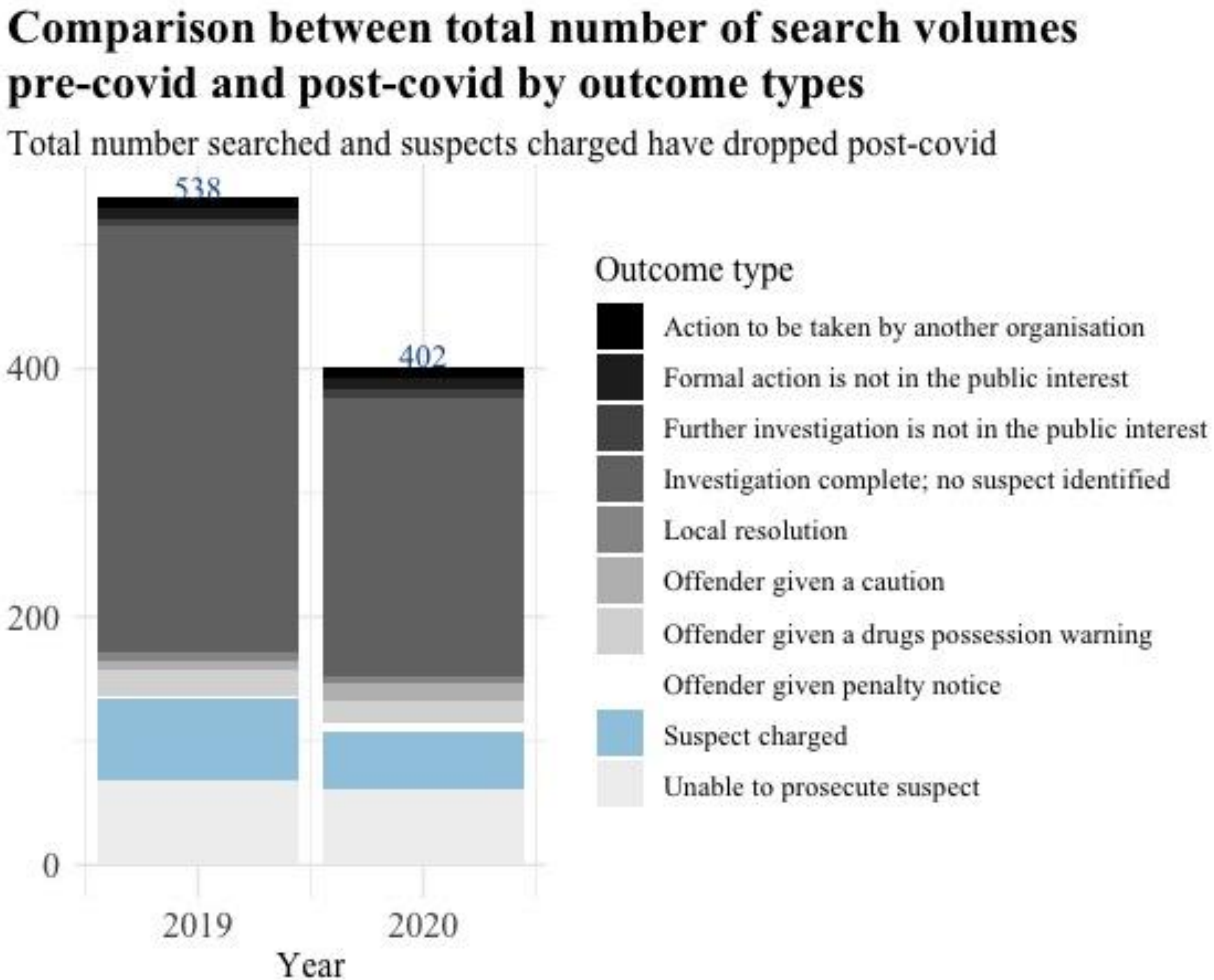


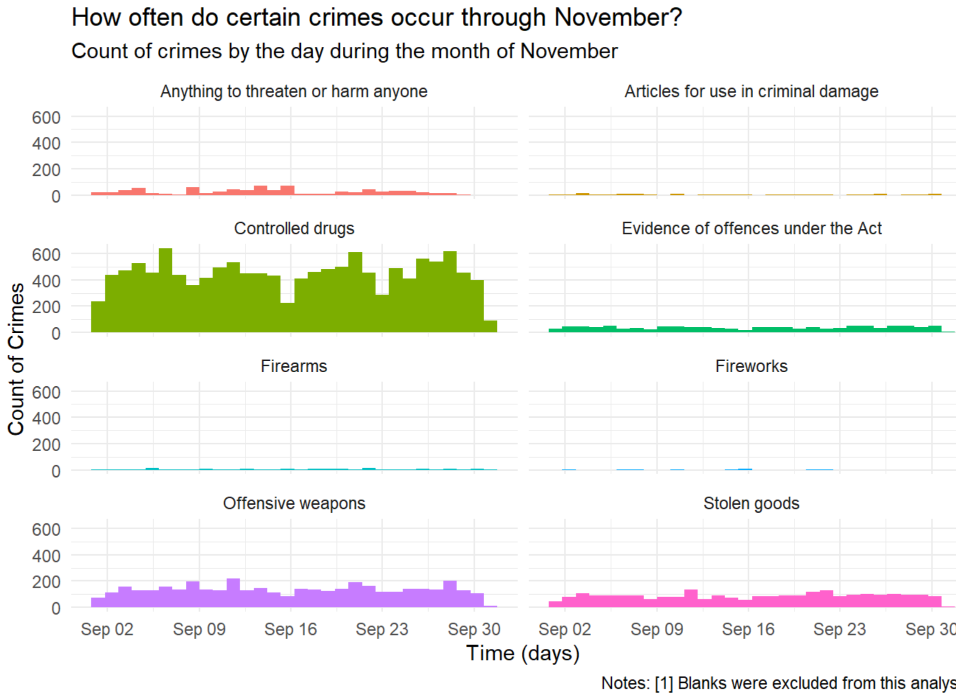


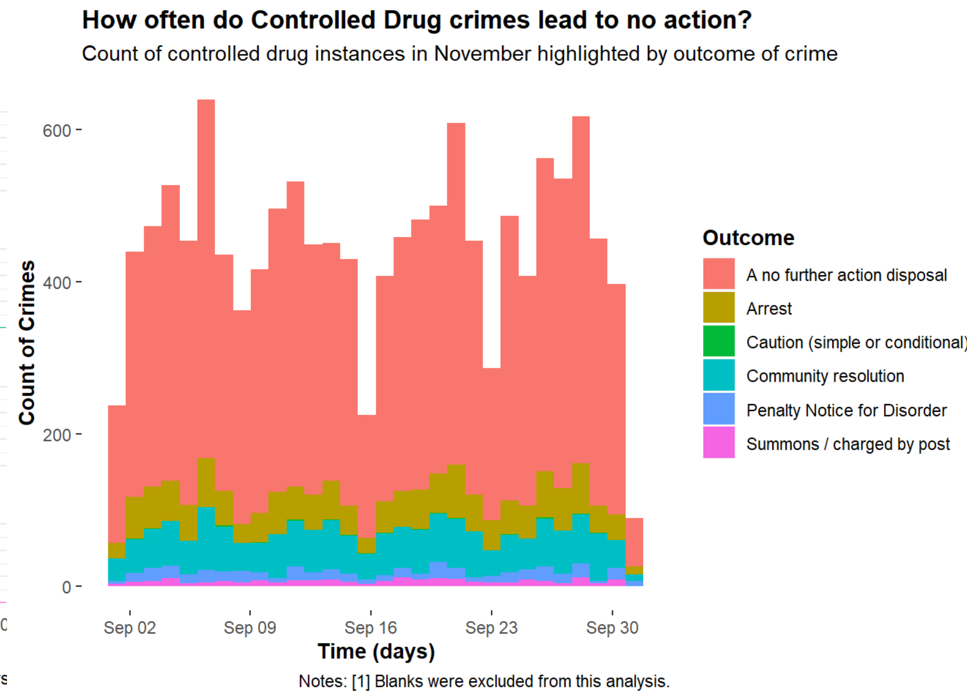

Sins of Youth pay a Price!
After leaving High School late teenagers/early adults are the most in trouble
Age Range
18-24
25-34
over 34
10-17
under 10
6634
33%
4217
21%
3502
18%
2947
15%
8
0%
Percentage of the total Stops and Searches
0
2000
4000
6000
Number of Stops and Searches in September 2020

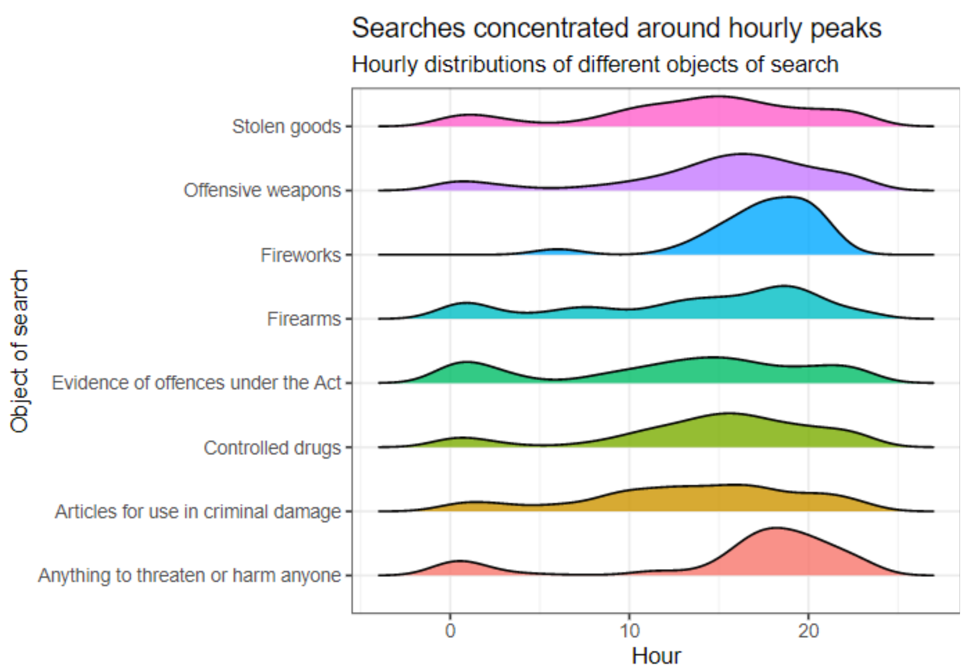

Drugs take the most attention!
Drugs related Stops and Searches are 4 times more frequent than the next Object of Search
Controlled drugs
Offensive weapons
Stolen goods
Evidence of offences under the Act
Anything to threaten or harm anyone
Articles for use in criminal damage
Firearms
Fireworks
12839
65%
3331
17%
2405
12%
842
4%
180
1%
168
1%
118
1%
7
0% Percentage of all Stops and Searches
Control of Drugs
take more than 65%
of the total Stops and Searches
0
5000
10000
Number of Stops and Searches in September 2020
Source: Data.Police.UK

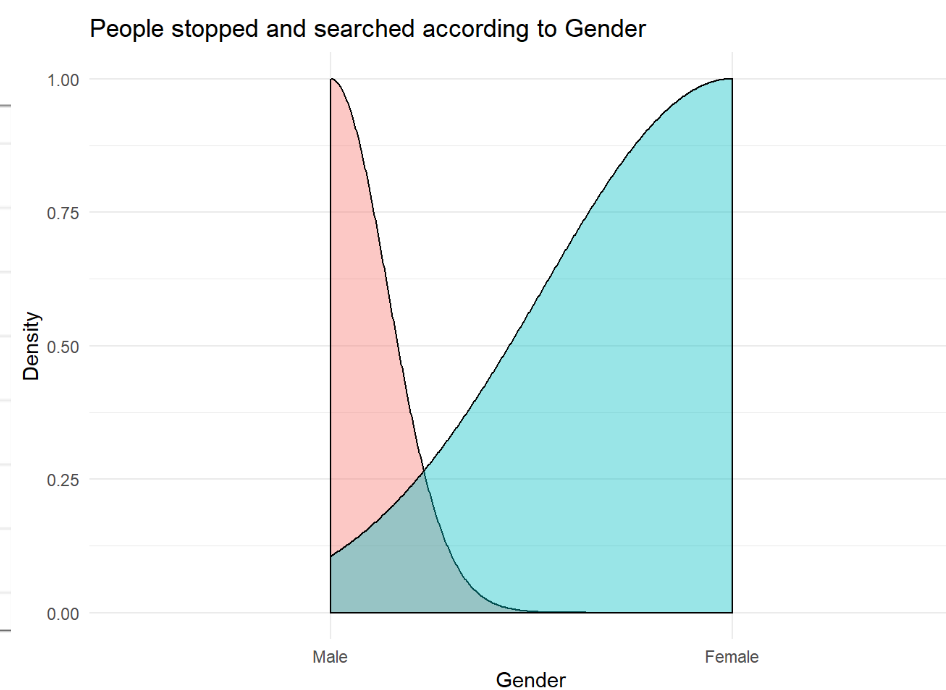
Multiple Distributions




How often do certain crimes occur through November?
Count of crimes by the day during the month of November
Anything to threaten or harm anyone
Articles for use in criminal damage
Controlled drugs
Evidence of offences under the Act
Firearms
Fireworks
Offensive weapons
Stolen goods
Count of Crimes
0
200
400
600
Sep 02
Sep 09
Sep 16
Sep 23
Sep 30
Time (days)
Notes: [1] Blanks were excluded from this analys

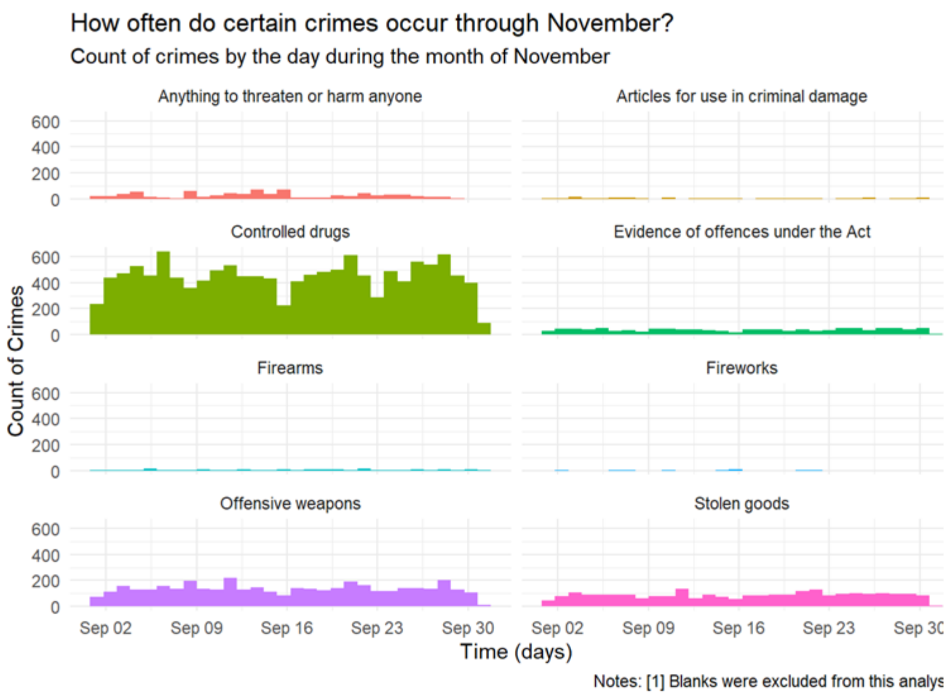

How often do Controlled Drug crimes lead to no action?
Count of controlled drug instances in November highlighted by outcome of crime
Count of Crimes
0
200
400
600
Sep 02
Sep 09
Sep 16
Sep 23
Sep 30
Time (days)
Outcome
A no further action disposal
Arrest
Caution (simple or conditional)
Community resolution
Penalty Notice for Disorder
Summons / charged by post
Notes: [1] Blanks were excluded from this analysis.


Searches concentrated around hourly peaks
Hourly distributions of different objects of search
Stolen goods
Offensive weapons
Fireworks
Firearms
Evidence of offences under the Act
Controlled drugs
Articles for use in criminal damage
Anything to threaten or harm anyone
Object of search
0
10
20
Hour

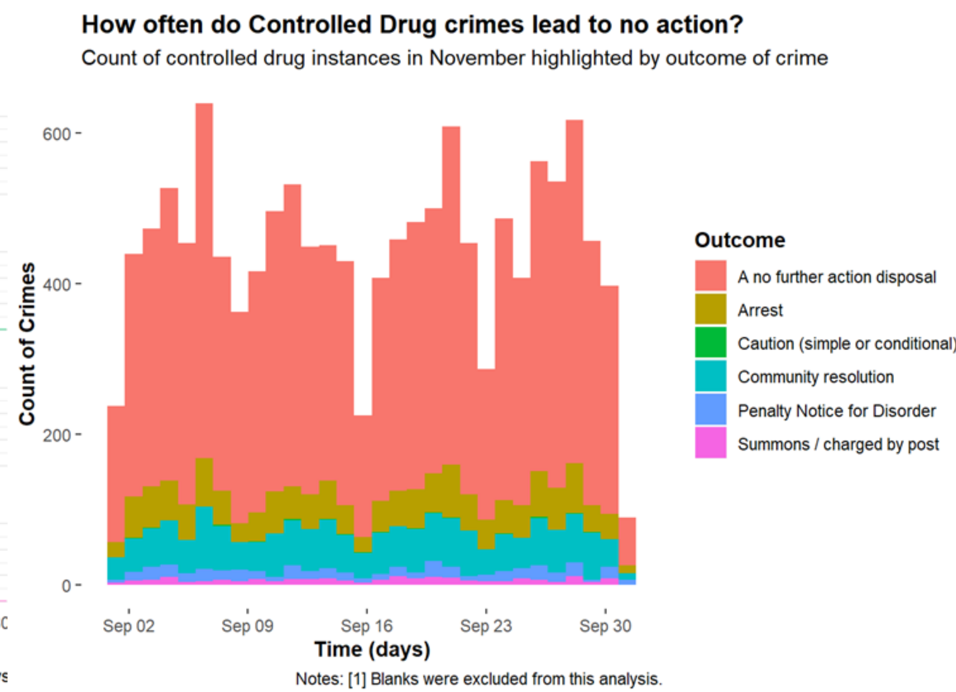

People stopped and searched according to Gender
Density
0.00
0.25
0.50
0.75
1.00
Male
Female
Gender

Multiple Distributions
`fill` with a different variable
`00_multiple_distributions.R`

OK, but perhaps remove empty categories? Or `scales = free`



Really hard to read

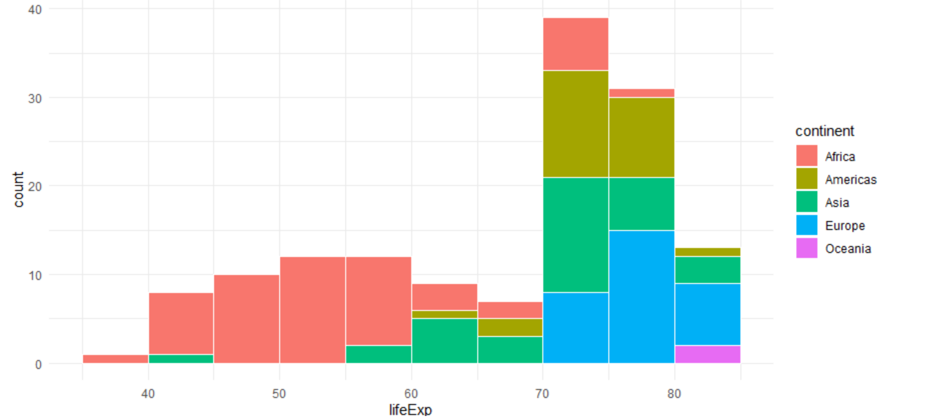


London Business School

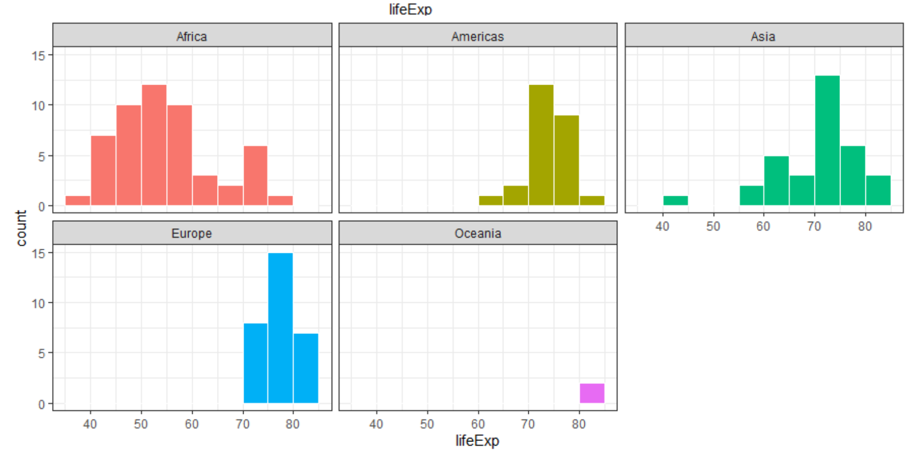
Multiple Distributions

00_multiple_distributions.R

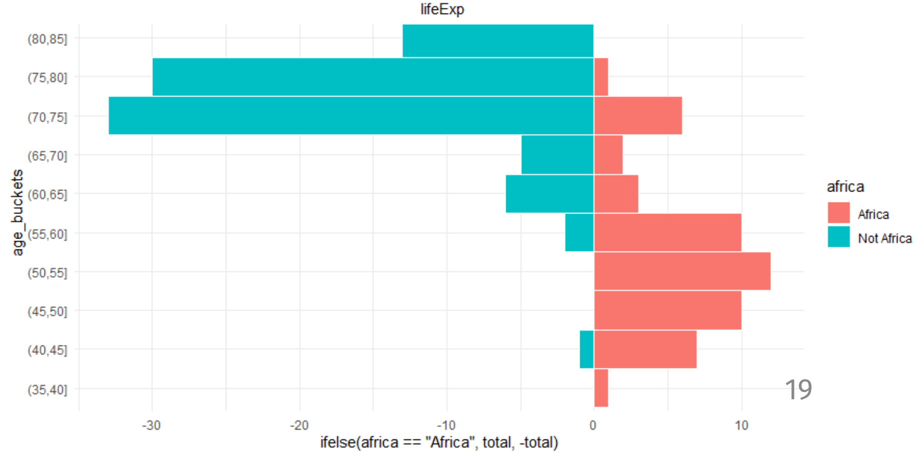
```r
# This makes it very hard to read and understand
ggplot(gapminder2007,
       aes(x = lifeExp,
           fill = continent)) +
  geom_histogram(binwidth = 5,
                 color = "white",
                 boundary = 50)+
  theme_minimal()+
  NULL
```




```r
# Facetting by continent makes it easier to read and understand
ggplot(gapminder2007,
       aes(x = lifeExp,
           fill = continent)) +
  geom_histogram(binwidth = 5,
                 color = "white",
                 boundary = 50)+
  guides(fill=FALSE)+
  theme_bw()+
  facet_wrap(~continent)+
  NULL
```



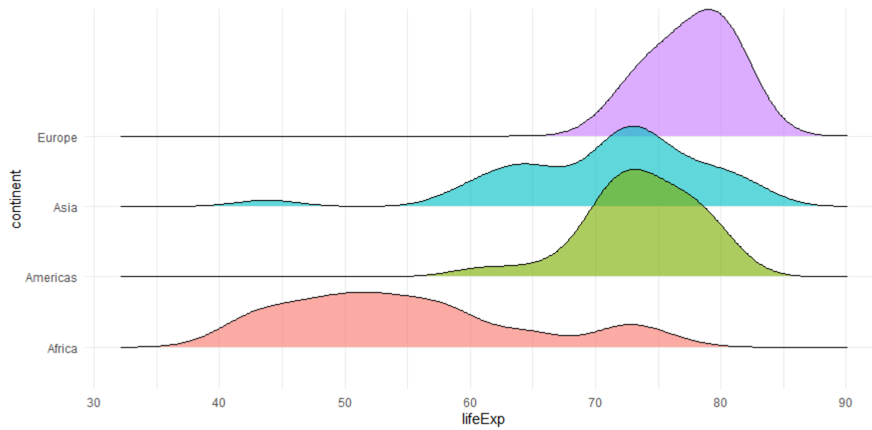

```r
# pyramid histograms
gapminder_intervals <- gapminder2007 %>%
  mutate(africa =
           ifelse(continent == "Africa",
                  "Africa", "Not Africa")) %>%
  mutate(age_buckets =
           cut(lifeExp,
               breaks = seq(30, 90, by = 5))) %>%
  group_by(africa, age_buckets) %>%
  summarize(total = n())

ggplot(gapminder_intervals,
       aes(y = age_buckets,
           x = ifelse(africa == "Africa",
                      total, -total),
           fill = africa)) +
  geom_col(width = 1, color = "white")+
  theme_minimal()+
  NULL
```



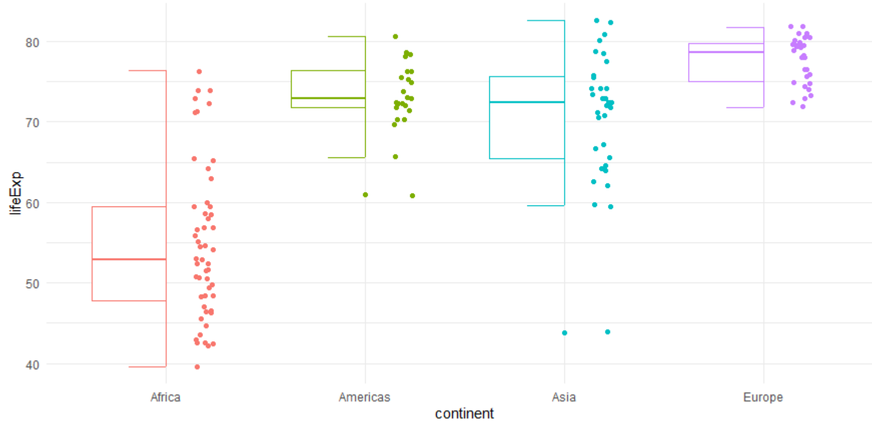
Multiple Distributions

00_multiple_distributions.R

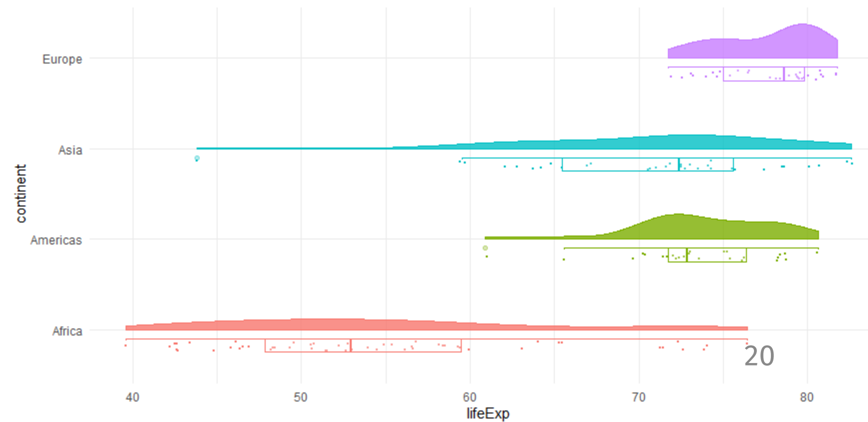
```
# use ggridges::geom_density_ridges() for multiple density plots
ggplot(filter(gapminder2007,
              continent != "Oceania"),
       aes(x = lifeExp,
           fill = continent,
           y = continent)) +
  geom_density_ridges(alpha = 5/8)+
  theme_minimal()+
  guides(fill=FALSE)+
  NULL
```




```
# use gghalves::geom_half_boxplot(), gghalves::geom_half_point()
ggplot(filter(gapminder2007,
              continent != "Oceania"),
       aes(y = lifeExp,
           x = continent,
           colour = continent)) +
  geom_half_boxplot(side = "l") + # half boxplot to the left
  geom_half_point(side = "r")+     # points to the right
  theme_minimal()+
  guides(fill = FALSE, color = FALSE)+
  NULL
```



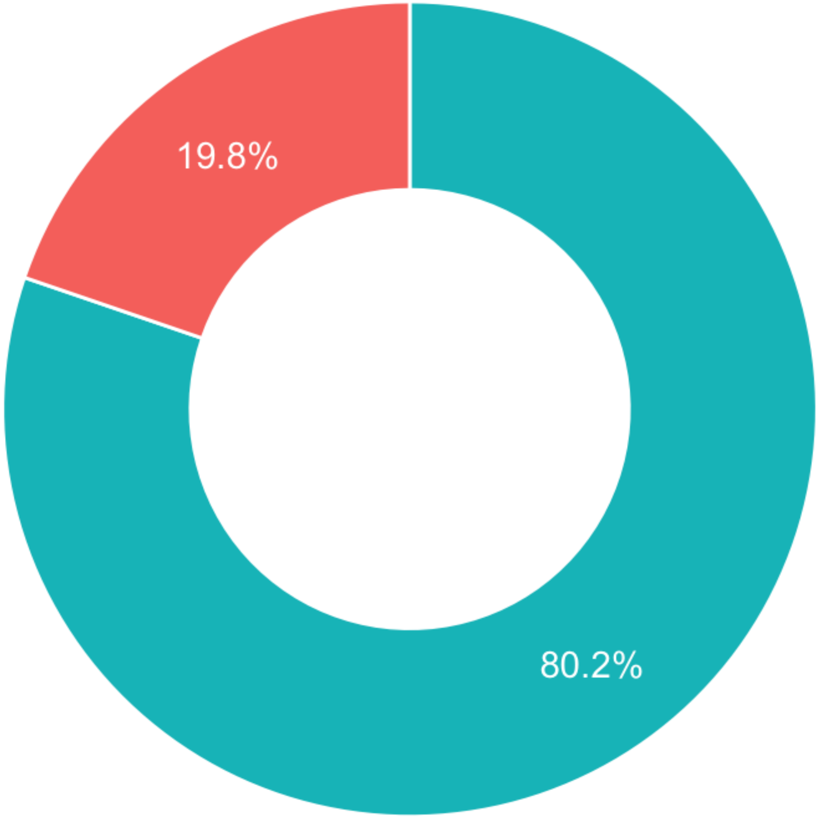

```
# Raincloud plots
ggplot(filter(gapminder2007,
              continent != "Oceania"),
       aes(y = lifeExp,
           x = continent,
           colour = continent)) +
  geom_half_point(side = "l", size = 0.3) +
  geom_half_boxplot(side = "l", width = 0.5,
                    alpha = 0.3, nudge = 0.1) +
  geom_half_violin(aes(fill = continent),
                   alpha = 0.8,
                   side = "r") +
  guides(fill = FALSE, color = FALSE) +
  coord_flip()+
  theme_minimal()+
  NULL
```



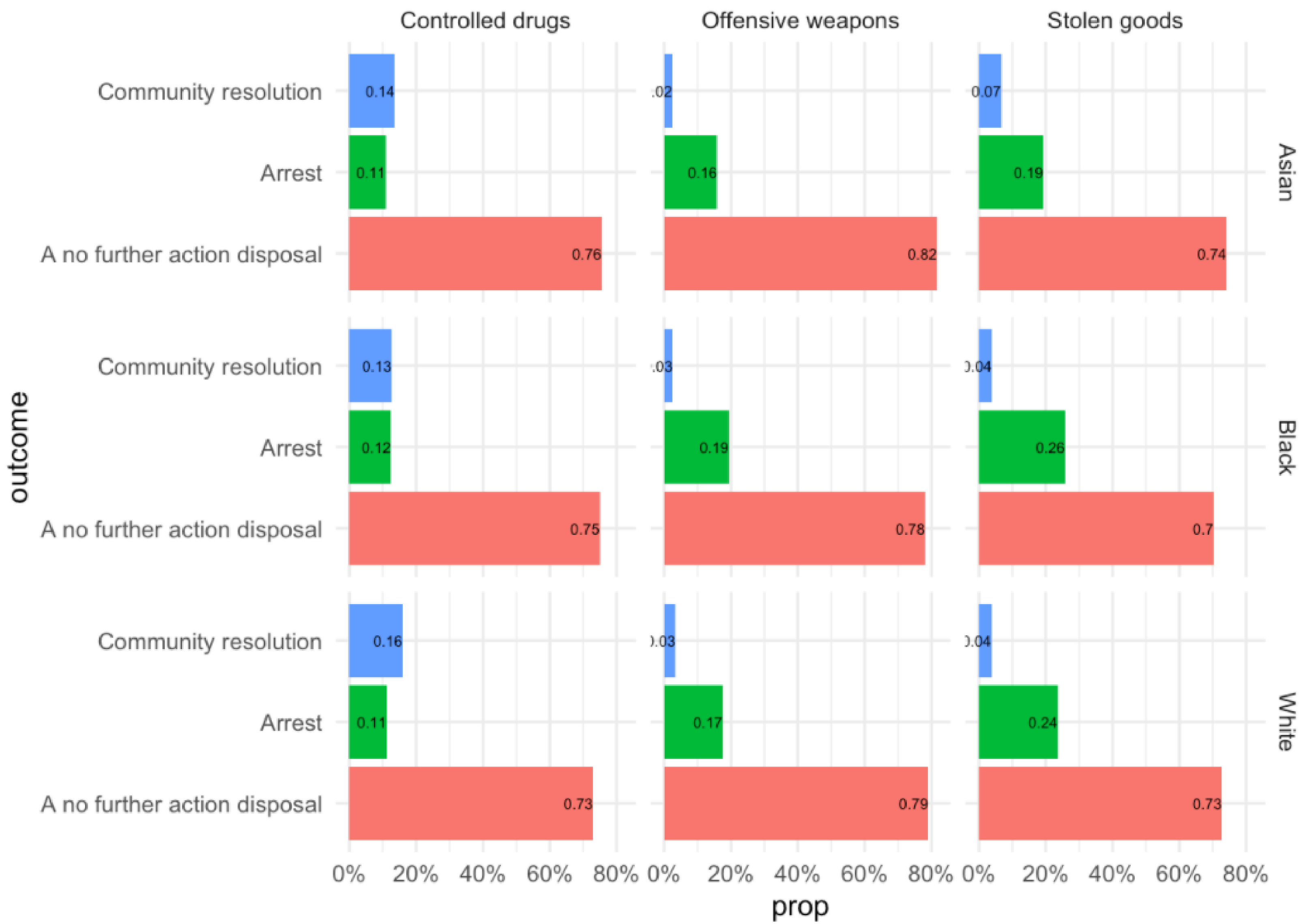
London Business School

How judgemental are the police when they decide your race?

Our analysis showed that almost one-fifth of police stops result in incorrect racial profiling



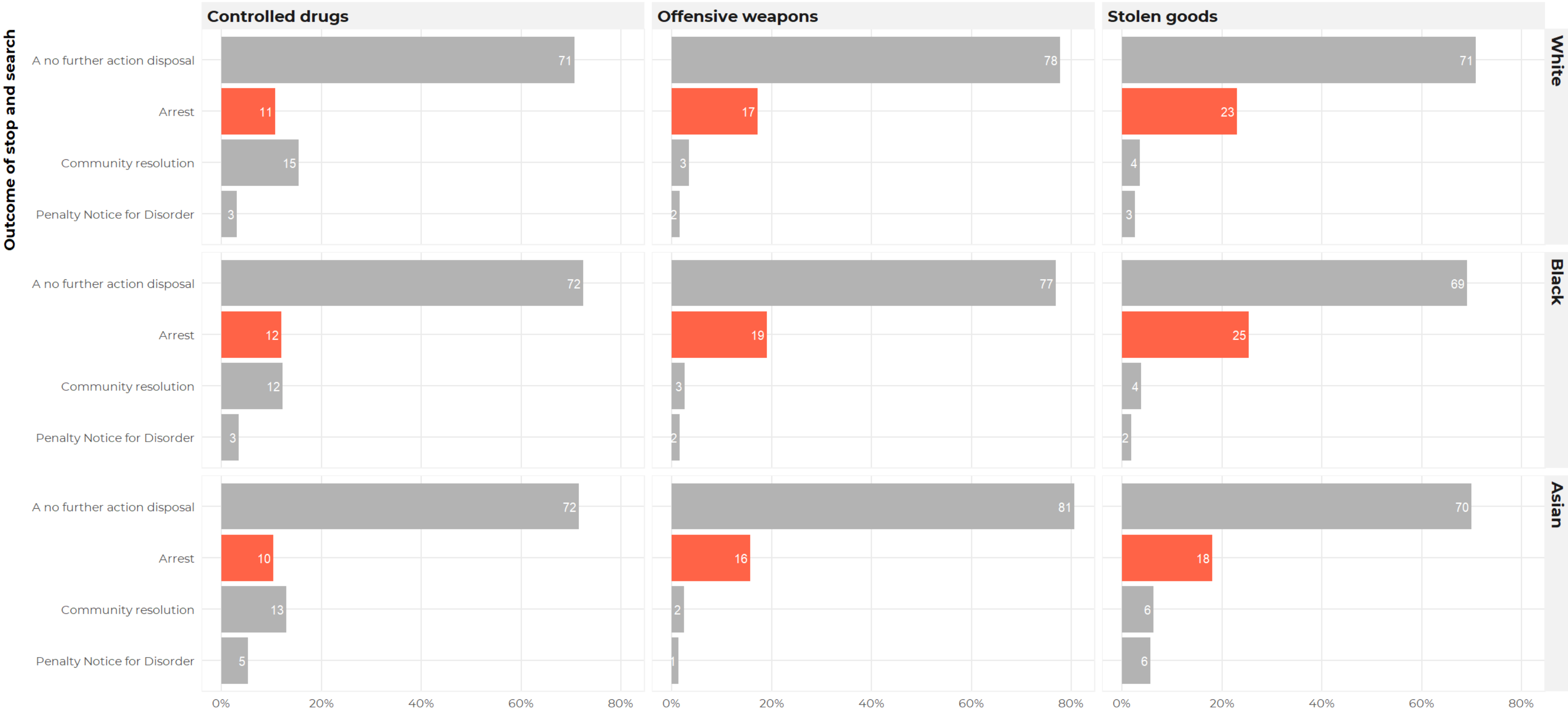
London Business School



London Business School

Blacks consistently have a higher % of arrests

Sep 2021



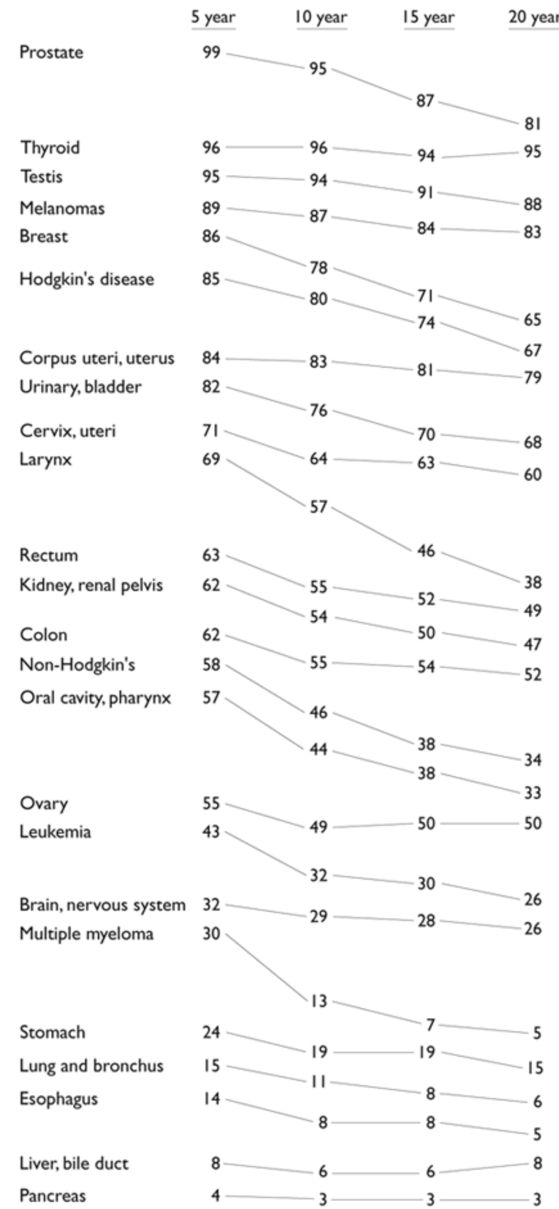
Stop & Search outcome by ethnicity and object of search

Source: Metropolitan Police

Slope graphs

Estimates of relative survival rates, by cancer site

	% survival rates and their standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8



https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000Jr