

Does race affect the result of traffic stops

Data

This dataset contains yearly traffic stops information provided by Charlotte-Mecklenburg Police Department(CMPD) available [here](#).

```
# Loading libraries needed
library(tidyverse)
library(knitr)
library(magrittr)
```

Loading the the Officer_Traffic_Stops.csv data

```
trfic_stps <- read.csv("Officer_Traffic_Stops.csv")

# taking unwanted columns out
trfic_stps <- trfic_stps[, !names(trfic_stps) %in%
                          c("ObjectID", "CreationDate", "Creator", "EditDate",
                            "Editor")]
```

```
str(trfic_stps)
```

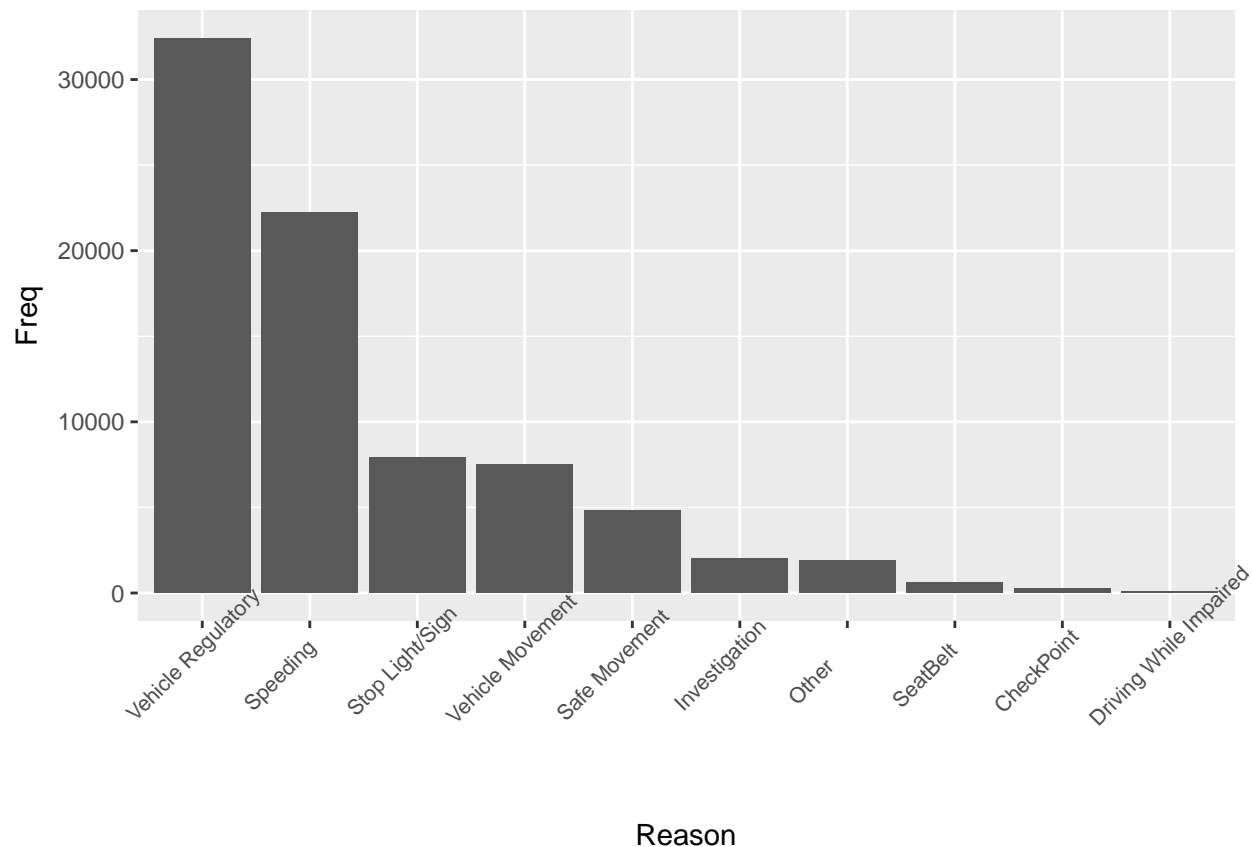
```
## 'data.frame': 79884 obs. of 12 variables:
## $ Month_of_Stop : Factor w/ 12 levels "2016/01","2016/02",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Reason_for_Stop : Factor w/ 10 levels "CheckPoint",...: 7 8 7 10 10 8 8 10 ...
## $ Officer_Race : Factor w/ 8 levels " ", "American Indian/Alaska Native",...: 8 8 8 4 8 8 ...
## $ Officer_Gender : Factor w/ 2 levels "Female", "Male": 2 2 2 2 2 2 2 1 2 ...
## $ Officer_Years_of_Service: int 6 6 6 2 6 6 6 2 7 9 ...
## $ Driver_Race : Factor w/ 5 levels "Asian", "Black",...: 5 2 2 2 5 5 5 2 2 5 ...
## $ Driver_Ethnicity : Factor w/ 2 levels "Hispanic", "Non-Hispanic": 2 2 2 2 2 2 1 2 2 2 ...
## $ Driver_Gender : Factor w/ 2 levels "Female", "Male": 2 2 2 1 2 1 2 2 1 2 ...
## $ Driver_Age : int 63 35 30 29 45 65 40 28 57 29 ...
## $ Was_a_Search_Conducted : Factor w/ 2 levels "No", "Yes": 1 1 1 1 1 1 1 1 1 2 ...
## $ Result_of_Stop : Factor w/ 5 levels "Arrest", "Citation Issued",...: 2 4 2 4 2 2 2 4 2 1 ...
## $ CMPD_Division : Factor w/ 14 levels " ", "Central Division",...: 3 3 3 7 3 3 5 7 7 14 ...
```

Dataframe trfic_stps has 79884 rows and 12 variables.

Lets see the reason for stopping vehicles by CMPD in year 2016

```
# count of reasons
stops <- as.data.frame(table(trfic_stps$Reason_for_Stop))
top3rsns <- as.character(arrange(stops, desc(Freq))$Var1[1:3])

# bar plot of reasons
stops %>% ggplot(aes(x = reorder(Var1, desc(Freq)), y = Freq)) +
  geom_bar(stat = "Identity") +
  theme(axis.text.x = element_text(angle=45, size = 8)) + xlab("Reason")
```



It can be seen from the above plot that main 3 causes for vehicle stopping by CPMD is due to violations in vehicle registration, speeding and Stop light/sign.

Lets understand actions taken by CMPD for each of reason of stopping.

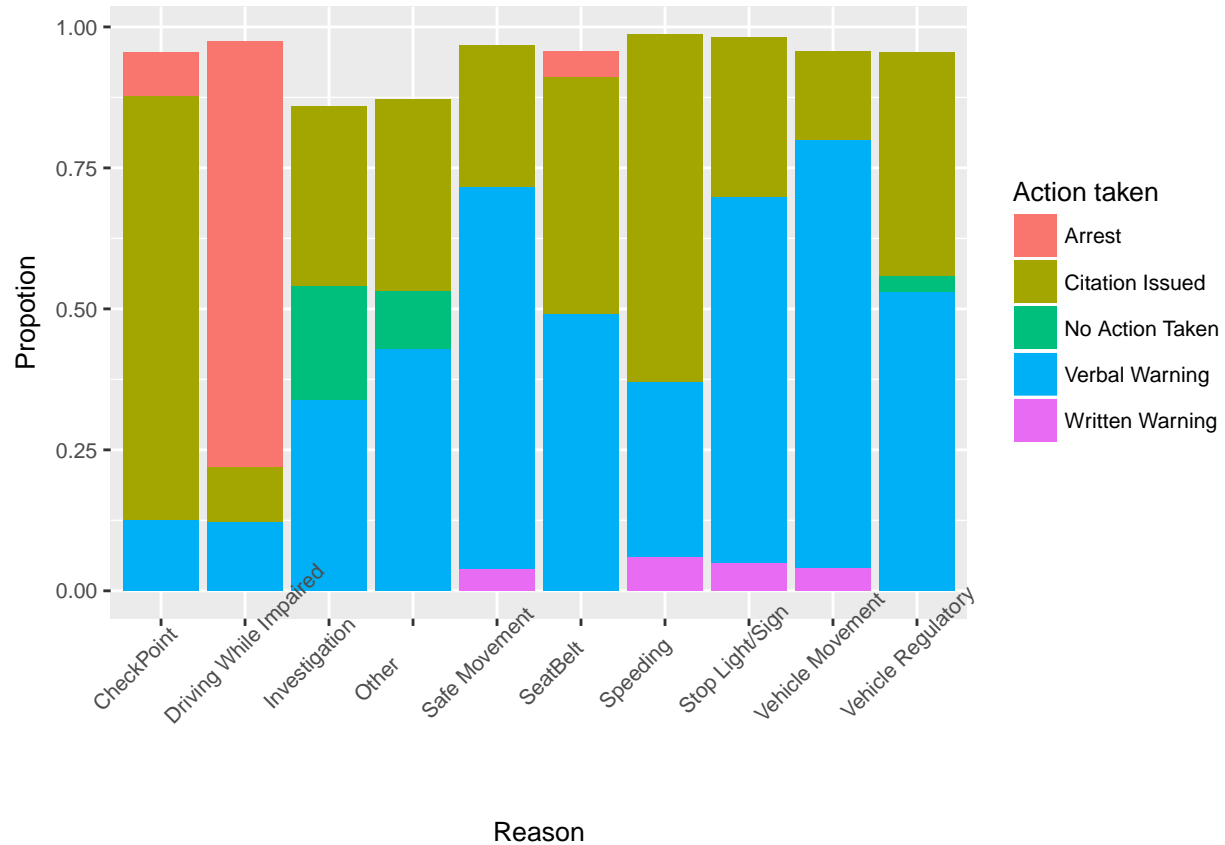
We will focus on top 3 actions taken by CMPD for each stopping event

```
# table with reason and result
consistncy <- as.data.frame(table(trfic_stps$Reason_for_Stop,
                                trfic_stps$Result_of_Stop))

# taking only first 3 results
consistncy %<>% group_by(Var1) %>% mutate(prop = Freq/sum(Freq)) %>%
  arrange(Var1, desc(Freq)) %>% slice(1:3)

# top 3 actions
top3rsn_rslt <- c(0,0,0)
for(i in 1:3)
  top3rsn_rslt[i] <- filter(consistncy, Var1 == top3rsns[i])[1,2]
top3rsn_rslt <- as.character(unlist(top3rsn_rslt))

ggplot(consistncy, aes(x = Var1, y = prop, fill = Var2)) +
  geom_bar(stat = "Identity") +
  theme(text = element_text(size=10),axis.text.x = element_text(angle=90)) +
  ylab("Propotion") + xlab("Reason") +
  scale_fill_discrete(name = "Action taken") +
  theme(axis.text.x = element_text(angle=45, size = 8))
```



Looking at the graph above it seems that there are quite bit variation in the action taken by CPMD. Still it can argued that action taken depends on the situation. Therefore it is worth looking at the profile of the driver.

However, before moving forward I am going to assume that CMPD take following actions regarding top 3 traffic violations in general. This will be the baseline.

```
kable(data.frame(Reson = top3rsns, General_result = top3rsn_rslt),caption = 'General action taken for t
```

Table 1: General action taken for top 3 stopping events

Reson	General_result
Vehicle Regulatory	Verbal Warning
Speeding	Citation Issued
Stop Light/Sign	Verbal Warning

Lets see whether driver profile (Race, Gender, Age) had any impact on the action taken by CPMD

```
# filter out the data for only top3 reasons
# create variable to identify consistent (0), and inconsisten (1) actions

new_trfic_stps <- trfic_stps %>% filter(Reason_for_Stop %in% top3rsns) %>%
  mutate(consistent = ifelse( (
    (Reason_for_Stop == top3rsns[1] & Result_of_Stop == top3rsn_rslt[1]) |
```

```

      (Reason_for_Stop == top3rsns[2] & Result_of_Stop == top3rsn_rslt[2]) |
      (Reason_for_Stop == top3rsns[3] & Result_of_Stop == top3rsn_rslt[3])
    ), yes = 0, no = 1))

# general rate being consistent
consisrate <- mean(new_trfic_stps$consistent == 0)

# for gender
consisrate_gender <- new_trfic_stps %>% group_by(Driver_Gender) %>%
  summarise(rate = mean(consistent == 0))

tdf1 <- data.frame(Variable = rep("Gender", nrow(consisrate_gender)),
                  Value = unlist(consisrate_gender[,2]),
                  Profile = unlist(consisrate_gender[,1]))

# for Driver_Race
consisrate_race <- new_trfic_stps %>% group_by(Driver_Race) %>%
  summarise(rate = mean(consistent == 0))

tdf2 <- data.frame(Variable = rep("Race", nrow(consisrate_race)),
                  Value = unlist(consisrate_race[,2]),
                  Profile = unlist(consisrate_race[,1]))

# creating gender groups
new_trfic_stps <- cbind(new_trfic_stps, Age_group =
  cut(new_trfic_stps$Driver_Age,
      breaks = c(0, 18, 55, 200),
      labels = c("Junior", "Mid Age", "Senior"),
      right = F))

consisrate_age <- new_trfic_stps %>% group_by(Age_group) %>%
  summarise(rate = mean(consistent == 0))

tdf3 <- data.frame(Variable = rep("Age_group", nrow(consisrate_age)),
                  Value = unlist(consisrate_age[,2]),
                  Profile = unlist(consisrate_age[,1]) )

# driver profiles
drivr_prof <- rbind(tdf1, tdf2, tdf3)

```

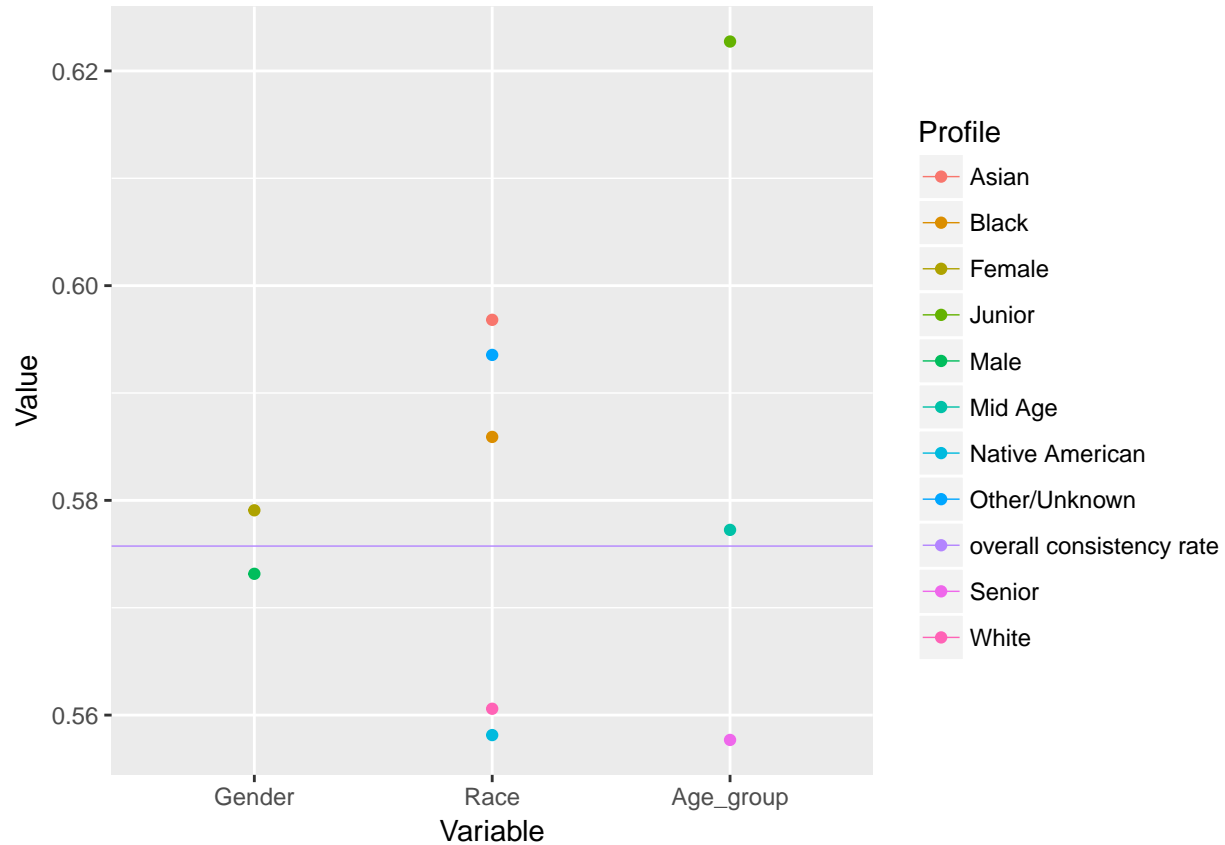
On top 3 traffic violations 58% of the time, action taken by CMPD is consistent with the general actions pointed out in Table 1. Now we are going to see how this consistency is affected with different driver profiles. So for e.g. if the calculated value is greater than 0.58, this means CMPD is more likely to take the general action for this particular driver group. **This will be identified as a consistent behavior.** And if the calculated value is less than 0.58, this means CMPD is comparatively less likely to take the general action for this particular driver group. **This will be identified as an inconsistent behavior**

```

# plot by driver profile
ggplot(drivr_prof, aes(x = Variable, y = Value, group = 1, col = Profile)) +
  geom_point() + geom_hline(aes(yintercept = consisrate,
                                col = "overall consistency rate"),
                           size=0.25, type = "b")

```

```
## Warning: Ignoring unknown parameters: type
```



It is visible from the above plot that actions taken by CMPD is more or less consistent for the gender of a driver. Relative to overall consistency rate, it seems that CMPD has been less consistent on races of white and native americans, where as they are more consistent on Asian and Black races. Also, relative to overall consistency rate, it seems that CMPD is more consistent on juniors where as they are less consistent on seniors. It will be interesting to know where these inconsistencies are leaning towards, i.e. whether CPMD being strict or lenient towards these groups.

On the other hand it will be interesting to look at profile of police officers to see whether it had a part to play in CMPD being less or more consistent.

Lets see whether how actions taken varies with the race of the driver and the profile of the police officer

Profile of the police officer is provided using attributes such as race, gender, years of service of the police officer.

```
# Race distribution of the driver
kable(count(x = new_trfic_stps, Driver_Race))
```

Driver_Race	n
Asian	1131
Black	33413
Native American	43
Other/Unknown	1860
White	26126

Table above shows the race distribution of the drivers. White and Black are the larger majority. Hence we will focus whether relatively higher consistency towards race of black we saw previously has anything to do with the profile of the police officer.

```
# filter only for black
new_trfic_stps_B <- filter(new_trfic_stps, Driver_Race == "Black")

# overall consistency rate for black
consisrate_B <- mean(new_trfic_stps_B$consistent == 0)

# by gender
consisrate_B_gender <- new_trfic_stps_B %>% group_by(Officer_Gender) %>%
  summarise(rate = mean(consistent == 0))

tdf4 <- data.frame(Variable = rep("Gender", nrow(consisrate_B_gender)),
  Value = unlist(consisrate_B_gender[,2]),
  Profile = unlist(consisrate_B_gender[,1]))

# by race
lvls <- as.character(levels(new_trfic_stps_B$Officer_Race)[2:8])
consisrate_B_race <- new_trfic_stps_B %>% filter(Officer_Race %in% lvls) %>%
  group_by(Officer_Race) %>%
  summarise(rate = mean(consistent == 0))

tdf5 <- data.frame(Variable = rep("Race", nrow(consisrate_B_race)),
  Value = unlist(consisrate_B_race[,2]),
  Profile = unlist(consisrate_B_race[,1]))

# by years of service
new_trfic_stps_B <- cbind(new_trfic_stps_B, Service_group =
  cut(new_trfic_stps_B$Officer_Years_of_Service,
    breaks = c(0, 3, 10, 100),
    labels = c("Newbie", "Mid career", "Experienced")))

consisrate_B_yos <- new_trfic_stps_B %>% group_by(Service_group) %>%
  summarise(rate = mean(consistent == 0))

tdf6 <- data.frame(Variable = rep("Service_group", nrow(consisrate_B_yos)),
  Value = unlist(consisrate_B_yos[,2]),
  Profile = unlist(consisrate_B_yos[,1]) )

# officer profiles
offcr_prof <- rbind(tdf4, tdf5, tdf6)
```

On overall consistency rate shown by CMPD is towards race of black is 0.59.

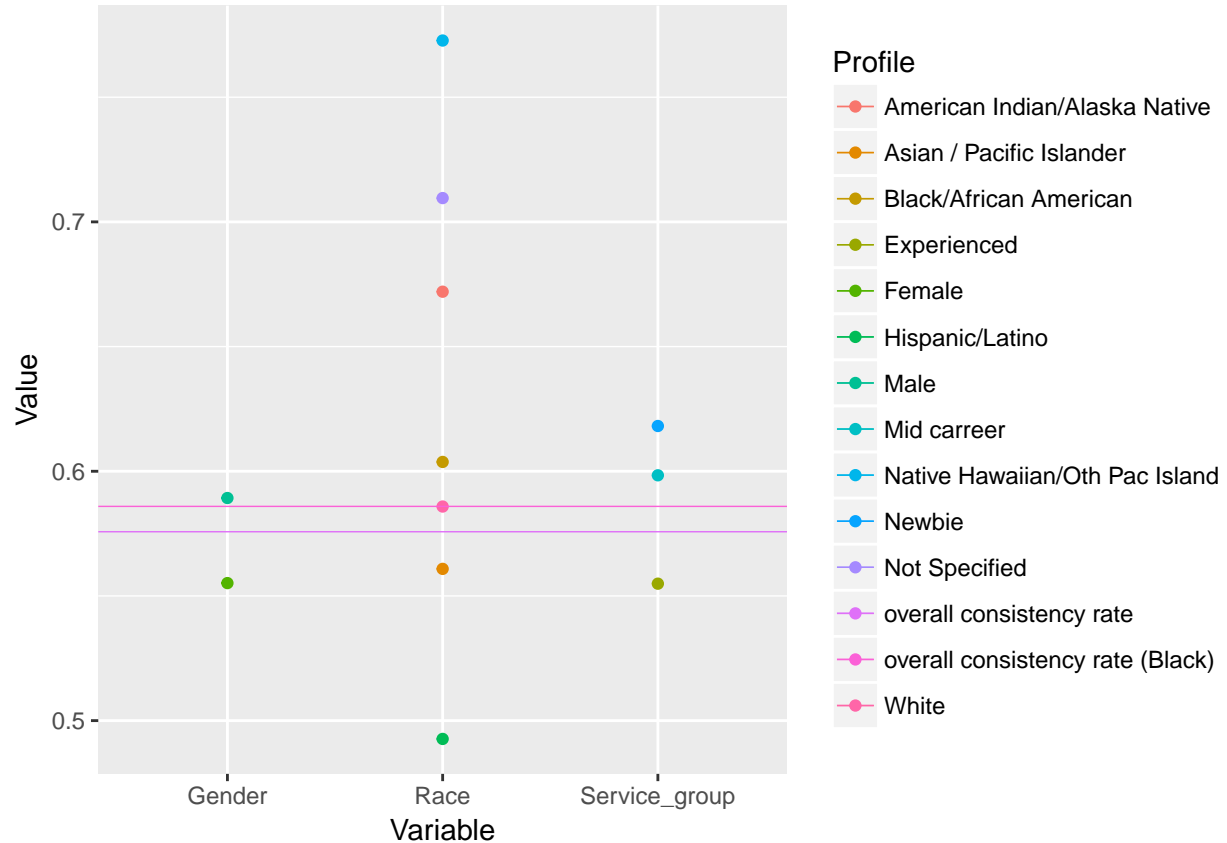
```
# plot by officer profile
offcr_prof$Profile <- as.character(offcr_prof$Profile)

ggplot(offcr_prof, aes(x = Variable, y = Value, group = 1, col = Profile)) +
  geom_point() + geom_hline(aes(yintercept = consisrate,
    col = "overall consistency rate"),
    size=0.25, type = "b") +
  geom_hline(aes(yintercept = consisrate_B,
    col = "overall consistency rate (Black)"),
```

```
size=0.25, type = "b")
```

```
## Warning: Ignoring unknown parameters: type
```

```
## Warning: Ignoring unknown parameters: type
```



Graph above indicate that female police officers tend to be less consistent in their action towards race of black. It can be seen that police officer with races from Native Hawaiian/Oth Pac Island, and American Indian/Alaska Native show more consistency towards race of black, whereas Hispanic/Latino police officers show less consistent approach. Also it can be seen that police officers who are new, shows more consistency towards race of black, where as experienced police officers have a less consistent approach. As pointed out previously it will interesting to look at where these inconsistencies are leaning towards, i.e. whether police officers being strict or lenient towards black groups.