Police_Data_Pei

General Model of all Data

Are police officers discrimatory?

```
glm.1 <- glm(search_conducted ~ driver_gender + driver_age + relevel(factor(data$driver_race), ref=4),
             data=data, family=binomial())
summary(glm.1)
```

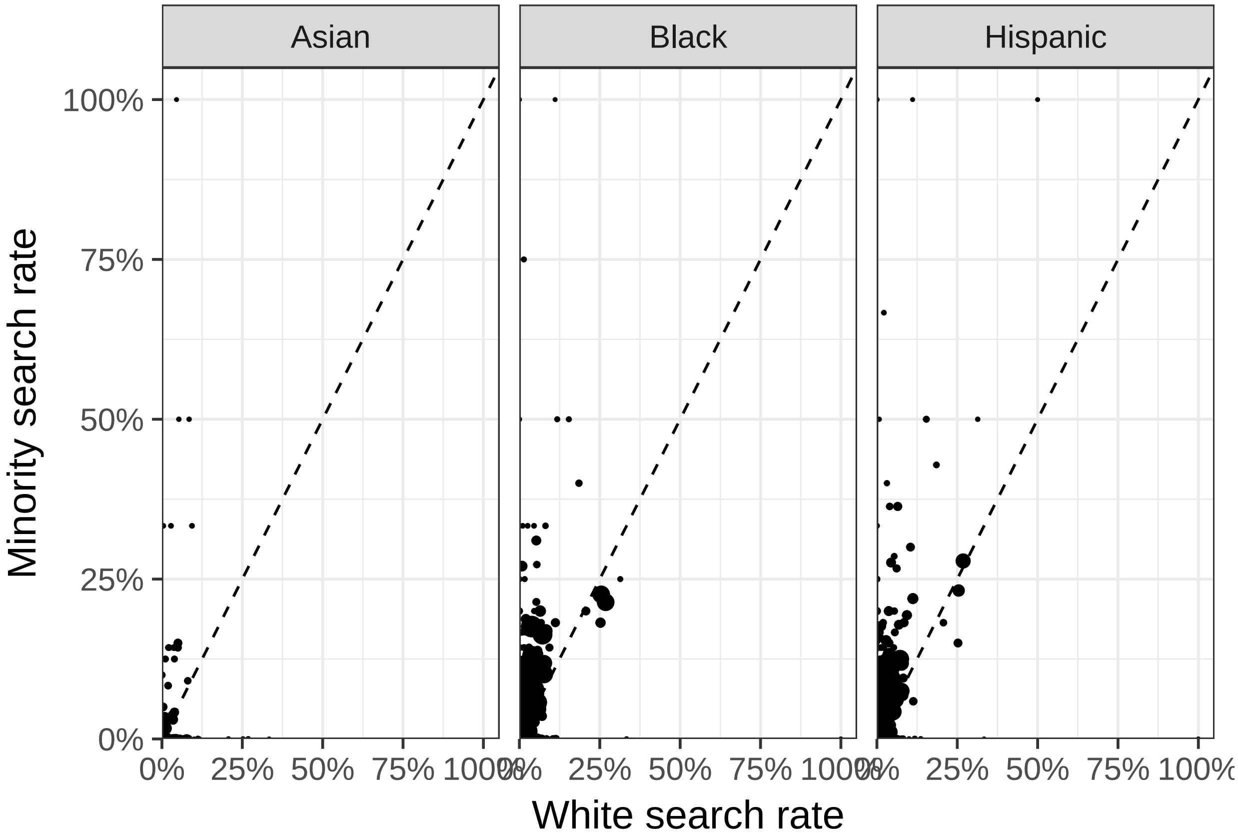
```
##
## Call:
## glm(formula = search_conducted ~ driver_gender + driver_age +
##     relevel(factor(data$driver_race), ref = 4), family = binomial(),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4468  -0.2198  -0.1475  -0.1035   4.0147
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                         -3.260939   0.053908
## driver_genderM                                       1.033257   0.038301
## driver_age                                          -0.054520   0.001349
## relevel(factor(data$driver_race), ref = 4)Asian     -1.169882   0.197615
## relevel(factor(data$driver_race), ref = 4)Black      0.845756   0.034506
## relevel(factor(data$driver_race), ref = 4)Hispanic   0.694880   0.037643
##                                                     z value Pr(>|z|)
## (Intercept)                                          -60.49  < 2e-16 ***
## driver_genderM                                        26.98  < 2e-16 ***
## driver_age                                           -40.42  < 2e-16 ***
## relevel(factor(data$driver_race), ref = 4)Asian       -5.92 3.22e-09 ***
## relevel(factor(data$driver_race), ref = 4)Black       24.51  < 2e-16 ***
## relevel(factor(data$driver_race), ref = 4)Hispanic    18.46  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 53972  on 316646  degrees of freedom
## Residual deviance: 50014  on 316641  degrees of freedom
##   (272 observations deleted due to missingness)
## AIC: 50026
##
## Number of Fisher Scoring iterations: 8
```

This general linear model says that being Black and Hispanics makes you more likely to be searched than if you were White (whereas if you were Asian, you would be less likely to be searched). Bar any other information, this makes it seem like the Conneticut Police are discrimatory against Blacks and Hispanics.

Let's take a closer look at the data.

Search Rate v Hit Rate: Officer Plots

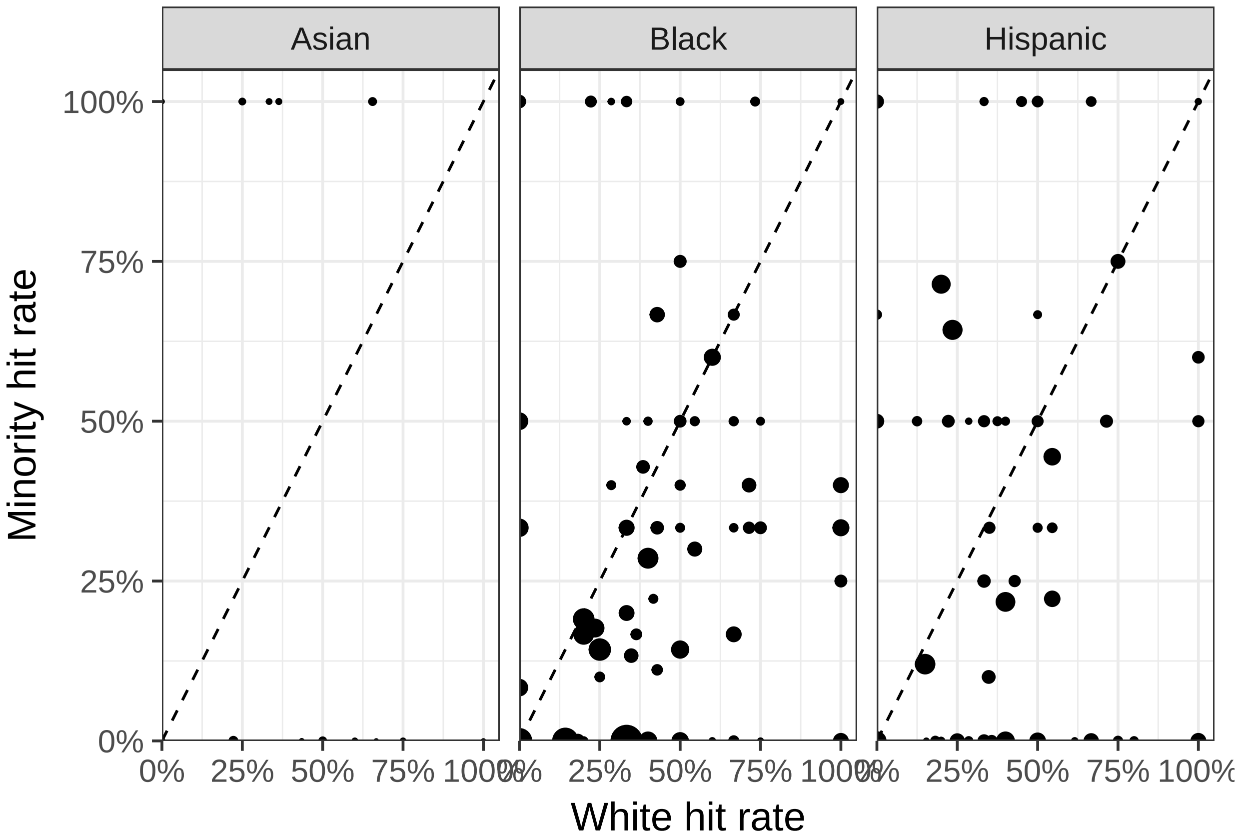
```
## Warning: Removed 32 rows containing missing values (geom_point).
```




```
## Warning: Unknown or uninitialised column: 'hitrate'.

## Warning in max(data_highsearchrate$hitrate): no non-missing arguments to
## max; returning -Inf

## Warning: Removed 97 rows containing missing values (geom_point).
```



The classic "outcome test" suggests that when searches of minority drivers are less likely to be successful, it may indicate that minority drivers are searched when less likely to be carrying contraband, suggesting discriminatory search standards. In general, the combination of higher search rates for minority drivers, along with lower hit rates, suggests minority drivers are being searched on less evidence.

In these plots, we plotted each individual police officer and their search rates and hit rates.

The first plot is of the white search rate for each officer, plotted against the minority search rate. Points that are above the diagonal line indicate that search rates are higher for minorities. While the search rate of most officers were definitely much higher for minorities, some officers had rates that were below the line too, and many where on the diaganol line.

Of corse, sometimes this disparity in white versus minority search rate could just be because of chance. So for the second plot, we took the police officers who's search rates were at least three times higher for minorities and plotted their hit rates.

```r
length(which(data_highsearchrate$hit_rate.y > data_highsearchrate$hit_rate.x &
               data_highsearchrate$driver_race.y == "Black"))
```

```
## [1] 37
```

```r
length(which(data_highsearchrate$hit_rate.y <= data_highsearchrate$hit_rate.x &
               data_highsearchrate$driver_race.y == "Black"))
```

```
## [1] 106
```

While there were indeed many officers, 37 of them, who had very high search rates for Black minorities and hit rates that are above the diagonal, so their hit rates were also higher for minorites, the majority of officers, 106 of them, who fit the inital search criteria that made them seem very discrimatory to begin with, have lower hit rates for minorities, suggesting discrimatory search standards for those officers.

Stop Outcomes

```
data_speeding <- data[which(data$violation == "Speeding"),]

multi.speeding_race <- multinom(stop_outcome ~ relevel(factor(data_speeding$driver_race), ref=4), data=d
summary(multi.speeding_race)
z_multi.speeding_race <- summary(multi.speeding_race)$coefficients/summary(multi.speeding_race)$standard
p_multi.speeding_race <- (1 - pnorm(abs(z_multi.speeding_race), 0, 1)) * 2
```

Let's consider another direction of determining discramatory practices. First, consider only the incidents where the stop was cited for Speeding reasons. Given that reason, is there discrimanation when the police officers give out a verbal warning, a written warning, a ticket, or a summons to court?

The mutinomial regression seems to indicate that there is White is the race that is most likely to get office with just a verbal warning, where as all other races have to deal with more several outcomes.

```
data_ticket_verbal <- data_speeding[which(data_speeding$stop_outcome == "Ticket" | data_speeding$stop_ou
data_ticket_verbal$stop_outcome <- factor(data_ticket_verbal$stop_outcome, levels=c("Verbal Warning", "T
data_ticket_verbal$driver_race <- factor(data_ticket_verbal$driver_race, levels=c("White", "Asian", "His
glm.ticket_verbal <- glm(stop_outcome ~ driver_race,
                         data=data_ticket_verbal,
                         family=binomial())
summary(glm.ticket_verbal)
```

```
##
## Call:
## glm(formula = stop_outcome ~ driver_race, family = binomial(),
##     data = data_ticket_verbal)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3135   0.4402   0.6063   0.6063   0.6063
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.60052    0.01059  151.17   <2e-16 ***
## driver_raceAsian     1.00437    0.08065   12.45   <2e-16 ***
## driver_raceHispanic  0.68493    0.03912   17.51   <2e-16 ***
## driver_raceBlack     0.38339    0.03079   12.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 73133  on 85962  degrees of freedom
## Residual deviance: 72498  on 85959  degrees of freedom
## AIC: 72506
##
## Number of Fisher Scoring iterations: 5
```

glm.ticket_verbal runs a regression on stop_outcome, where outcome is only either Ticket or Verbal Warning. Asians, Hispanics, and Blakcs are all more likely than Whites to recieve a ticket over getting off with a verbal warning, with coefficients of 1.00437, 0.68493, and 0.38339, respectively.

```
data_summons_verbal <- data_speeding[which(data_speeding$stop_outcome == "Summons" | data_speeding$stop_
data_summons_verbal$stop_outcome <- factor(data_summons_verbal$stop_outcome, levels=c("Verbal Warning",
```

```r
data_summons_verbal$driver_race <- factor(data_summons_verbal$driver_race, levels=c("White", "Asian", "
glm.summons_verbal <- glm(stop_outcome ~ driver_race,
                          data=data_summons_verbal,
                          family=binomial())
summary(glm.summons_verbal)
```

```
##
## Call:
## glm(formula = stop_outcome ~ driver_race, family = binomial(),
##     data = data_summons_verbal)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7905  -0.4739  -0.4739  -0.4739   2.2101
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -2.13004    0.02963 -71.880   <2e-16 ***
## driver_raceAsian    -0.22133    0.26323  -0.841      0.4
## driver_raceHispanic  1.12709    0.07533  14.963   <2e-16 ***
## driver_raceBlack     0.93752    0.06352  14.760   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11739  on 15014  degrees of freedom
## Residual deviance: 11390  on 15011  degrees of freedom
## AIC: 11398
##
## Number of Fisher Scoring iterations: 4
```

glm.summons_verbal runs a regression on stop_outcome, where outcome is only either Summons, the most severe outcome, or verbal warning. Hispanics and Blacks were more likely than whites to recieve summons, with coefficients of 1.12709 and 0.93752 respectively.