Aim1_traffic_violations

Cameron Baker

4/11/2021

```r
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
final_df <- read.csv("2019_Racial_Profiling_Combined.csv", stringsAsFactors = F)
moving_traffic_violations <-  filter(final_df, Reason_for_stop == "Moving Traffic Violation") %>% filte
moving_traffic_violations <- moving_traffic_violations[which(moving_traffic_violations$PRIMARY_KEY %in%
```

```r
pop_stats_2010 <- read.csv("2010 census data.csv")
pop_stats_2010[1,]
```

```
##                 NAME Total.pop Pop.One.Race Pop.Two.More.Races Black.pop
## 1 Austin city, Texas    614925       598291              16634     48230
##   Native.pop Asian.pop PI.pop White.pop Hispanic.pop
## 1       4807     39777    432    329500       188318
```

```r
Black.pop <- (sum(pop_stats_2010$Black.pop) / sum(pop_stats_2010$Total.pop))
Hispanic.pop <- (sum(pop_stats_2010$Hispanic.pop) / sum(pop_stats_2010$Total.pop))
White.pop <- (sum(pop_stats_2010$White.pop) / sum(pop_stats_2010$Total.pop))


table(moving_traffic_violations$type)
```

```
##
##   arrest citation  warning
##     6327    24067    67907
```

```r
table(moving_traffic_violations$type, moving_traffic_violations$Race)
```

```
##
##            BLACK HISPANIC OR LATINO WHITE
##   arrest    1805               2797  1725
##   citation  3102               8943 12022
##   warning   9686              19964 38257
```

```
traffic_stop_sex_tbl <- as.data.frame.matrix(table(moving_traffic_violations$type, moving_traffic_viola
traffic_stop_sex_tbl <- traffic_stop_sex_tbl[,c(2,3)]

#traffic_stop_sex_tbl[,1] <- round(traffic_stop_sex_tbl[,1] / sum(traffic_stop_sex_tbl[,1]),3)
#traffic_stop_sex_tbl[,2] <- round(traffic_stop_sex_tbl[,2] / sum(traffic_stop_sex_tbl[,2]),3)

knitr::kable(t(traffic_stop_sex_tbl))
```

	arrest	citation	warning
F	1577	9154	27781
M	4750	14909	40122

```
traffic_stop_sex_tbl <- as.data.frame.matrix(table(moving_traffic_violations$type, moving_traffic_viola
traffic_stop_sex_tbl <- traffic_stop_sex_tbl[,c(2,3)]

#traffic_stop_sex_tbl[1,] <- traffic_stop_sex_tbl[1,] / sum(traffic_stop_sex_tbl[1,]),3)
#traffic_stop_sex_tbl[2,] <- traffic_stop_sex_tbl[2,] / sum(traffic_stop_sex_tbl[2,]),3)
#traffic_stop_sex_tbl[3,] <- traffic_stop_sex_tbl[3,] / sum(traffic_stop_sex_tbl[3,]),3)

knitr::kable(t(traffic_stop_sex_tbl))
```

	arrest	citation	warning
F	1577	9154	27781
M	4750	14909	40122

```
traffic_stop_race_tbl <- t(as.data.frame.matrix(table(moving_traffic_violations$type, moving_traffic_vi

round(rowSums(traffic_stop_race_tbl) / sum(traffic_stop_race_tbl),4)
```

```
##              BLACK HISPANIC OR LATINO              WHITE
##             0.1485             0.3225             0.5290
```

```
knitr::kable(traffic_stop_race_tbl)
```

	arrest	citation	warning
BLACK	1805	3102	9686
HISPANIC OR LATINO	2797	8943	19964
WHITE	1725	12022	38257

```
traffic_stop_race_tbl[1,] <- round(traffic_stop_race_tbl[1,] / sum(traffic_stop_race_tbl[1,]),3)
traffic_stop_race_tbl[2,] <- round(traffic_stop_race_tbl[2,] / sum(traffic_stop_race_tbl[2,]),3)
traffic_stop_race_tbl[3,] <- round(traffic_stop_race_tbl[3,] / sum(traffic_stop_race_tbl[3,]),3)

knitr::kable(traffic_stop_race_tbl)
```

	arrest	citation	warning
BLACK	0.124	0.213	0.664
HISPANIC OR LATINO	0.088	0.282	0.630
WHITE	0.033	0.231	0.736

```r
moving_traffic_violations$Race_Sex <- paste0(moving_traffic_violations$Race," ",moving_traffic_violation
traffic_stop_race_tbl <- as.data.frame.matrix(table(moving_traffic_violations$type, moving_traffic_viola
traffic_stop_race_tbl <- traffic_stop_race_tbl[,c(2,3,5,6,9,10)]
#traffic_stop_race_tbl[,1] <- round(traffic_stop_race_tbl[,1] / sum(traffic_stop_race_tbl[,1]),3)
#traffic_stop_race_tbl[,2] <- round(traffic_stop_race_tbl[,2] / sum(traffic_stop_race_tbl[,2]),3)
#traffic_stop_race_tbl[,3] <- round(traffic_stop_race_tbl[,3] / sum(traffic_stop_race_tbl[,3]),3)
#traffic_stop_race_tbl[,4] <- round(traffic_stop_race_tbl[,4] / sum(traffic_stop_race_tbl[,4]),3)
#traffic_stop_race_tbl[,5] <- round(traffic_stop_race_tbl[,5] / sum(traffic_stop_race_tbl[,5]),3)
#traffic_stop_race_tbl[,6] <- round(traffic_stop_race_tbl[,6] / sum(traffic_stop_race_tbl[,6]),3)
traffic_stop_race_tbl <- t(traffic_stop_race_tbl)

knitr::kable(traffic_stop_race_tbl)
```
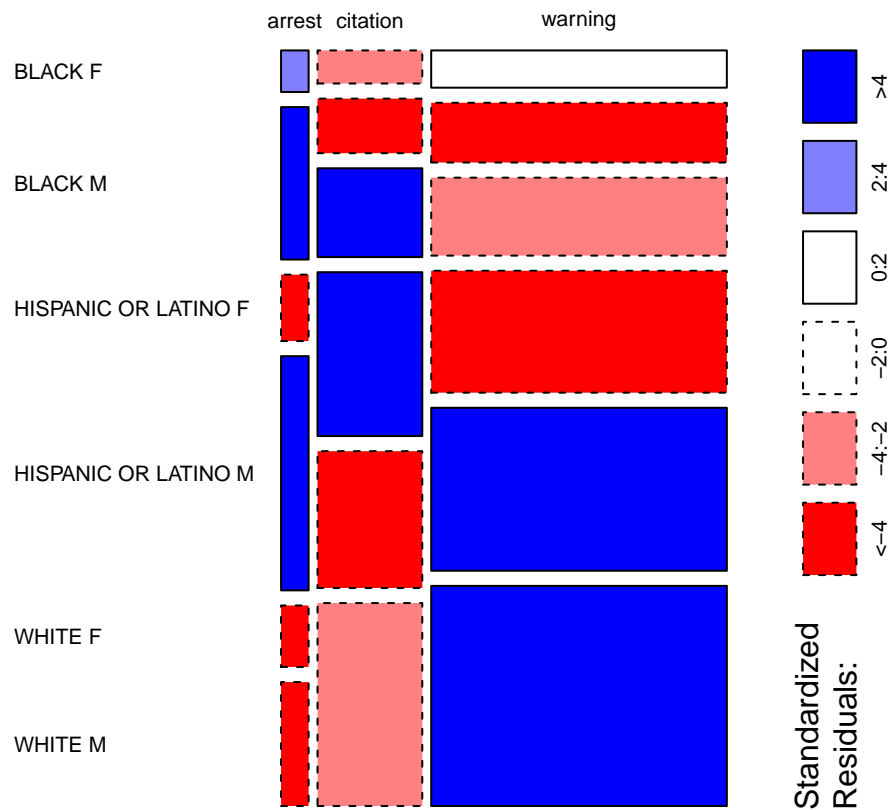
	arrest	citation	warning
BLACK F	387	1171	3720
BLACK M	1418	1930	5966
HISPANIC OR LATINO F	618	3143	7788
HISPANIC OR LATINO M	2179	5798	12175
WHITE F	572	4840	16273
WHITE M	1153	7181	21981

```r
library(graphics)
mosaicplot(t(traffic_stop_race_tbl), shade = TRUE, las=1, main = "")
```
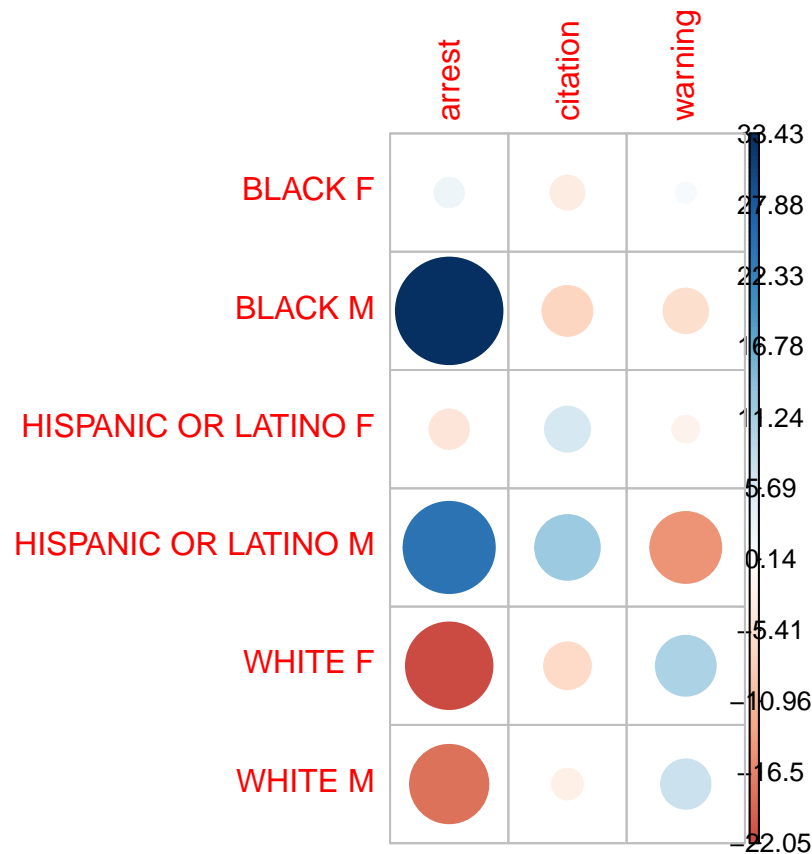
```
out_chisq <- chisq.test(t(traffic_stop_race_tbl))
out_chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  t(traffic_stop_race_tbl)
## X-squared = 3280.8, df = 10, p-value < 2.2e-16
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(t(out_chisq$residuals), is.cor = FALSE)
```

```r
#library(gplots)
#traffic_stop_race_tbl <- as.matrix(table(moving_traffic_violations$type, moving_traffic_violations$Rac
#traffic_stop_race_tbl <- as.table(traffic_stop_race_tbl[,c(2,3,5,6,9,10)])
#balloonplot(traffic_stop_race_tbl, label=F)
#library(reshape2)
#melted_tbl <- melt(traffic_stop_race_tbl)
#melted_tbl$outcome <- rep(c("Arrest","Citation","Warning"),3)
#ggplot(melted_tbl, aes(x = variable, y = value, color = outcome)) + geom_bar()
```

```r
moving_traffic_violations$parsed_time <- as.numeric(gsub(":.*$","",moving_traffic_violations$Time))
moving_traffic_violations$Time_of_day <- "Night"
moving_traffic_violations$Time_of_day[which(moving_traffic_violations$parsed_time > 6 & moving_traffic_v
knitr::kable(table(moving_traffic_violations$Time_of_day, moving_traffic_violations$type))
```

	arrest	citation	warning
Day	2385	21087	55354
Night	3942	2980	12553

```r
moving_traffic_violations$RACE_TOD <- paste0(moving_traffic_violations$Race," (",moving_traffic_violatio

race_tod_tbl <- table(moving_traffic_violations$RACE_TOD, moving_traffic_violations$type)

knitr::kable(race_tod_tbl)
```

	arrest	citation	warning
BLACK (Day)	835	2669	7253
BLACK (Night)	970	433	2433
HISPANIC OR LATINO (Day)	1012	7736	15928
HISPANIC OR LATINO (Night)	1785	1207	4036
WHITE (Day)	538	10682	32173
WHITE (Night)	1187	1340	6084

```
race_known_tbl <- table(moving_traffic_violations$RACE_TOD, moving_traffic_violations$Race_known)

tmp <- as.data.frame.matrix(race_known_tbl)
sum(tmp[,2]) / (sum(tmp[,1]) + sum(tmp[,2]))
```

```
## [1] 0.00832061
```

```
sum(c(tmp[1,2],tmp[3,2],tmp[5,2])) / (sum(tmp[1,]) + sum(tmp[3,]) + sum(tmp[5,]))
```

```
## [1] 0.009276722
```

```
night_true <- sum(c(tmp[2,2],tmp[4,2],tmp[6,2]))
night_total <- sum(tmp[2,]) + sum(tmp[4,]) + sum(tmp[6,])

prop.test(c(night_true,sum(tmp[,2])), c(night_total,sum(colSums(tmp))))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(night_true, sum(tmp[, 2])) out of c(night_total, sum(colSums(tmp)))
## X-squared = 31.817, df = 1, p-value = 1.694e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.005054359 -0.002797032
## sample estimates:
##      prop 1      prop 2
## 0.004394914 0.008320610
```
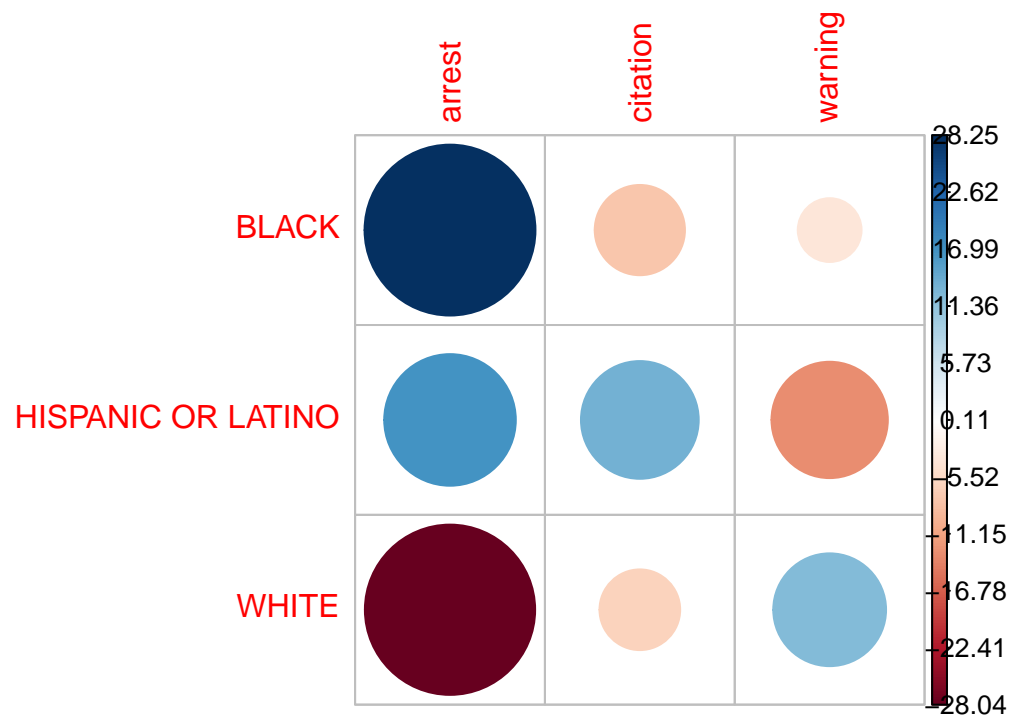
```
knitr::kable(race_known_tbl)
```

|	FALSE	TRUE
BLACK (Day)	10558	91
BLACK (Night)	3732	16
HISPANIC OR LATINO (Day)	24279	243
HISPANIC OR LATINO (Night)	6819	28
WHITE (Day)	42911	394
WHITE (Night)	8478	40

```
race.chisq <- chisq.test(table(moving_traffic_violations$Race, moving_traffic_violations$type))
corrplot(race.chisq$residuals, is.cor = FALSE)
```

```
race_tod_tbl.chisq <- chisq.test(race_tod_tbl)
corrplot(race_tod_tbl.chisq$residuals, is.cor = FALSE)
```
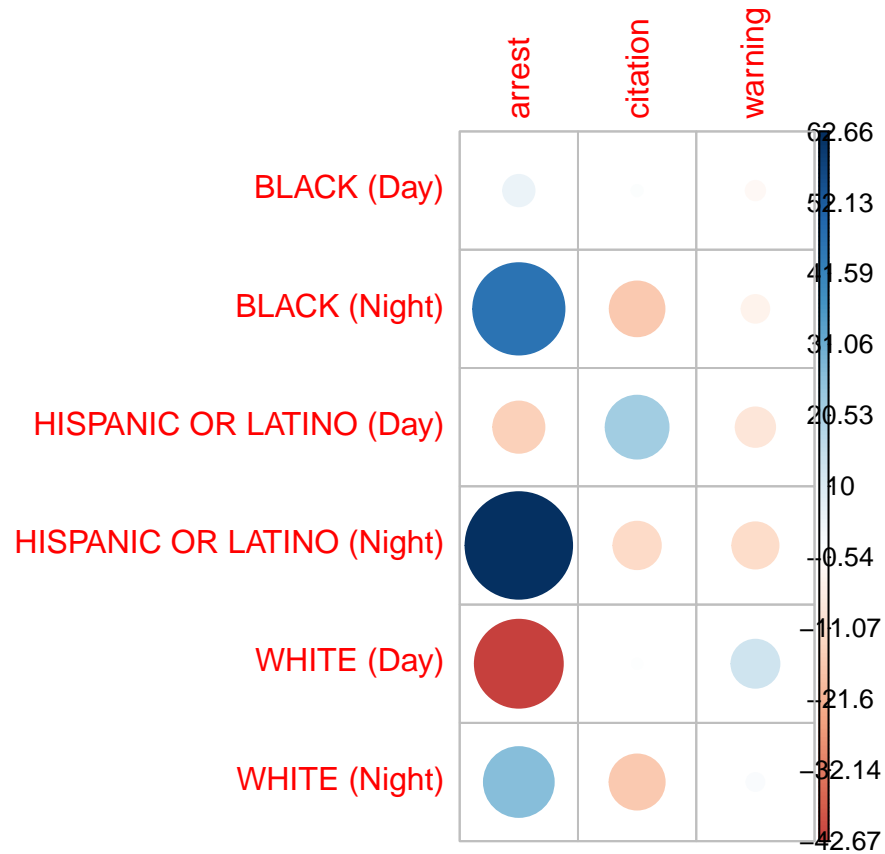
```
knitr::kable(table(moving_traffic_violations$Race_known,moving_traffic_violations$type))
```

	arrest	citation	warning
FALSE	5615	23324	67838
TRUE	0	743	69

```
knitr::kable(table(moving_traffic_violations$Race_known,moving_traffic_violations$Searched))
```

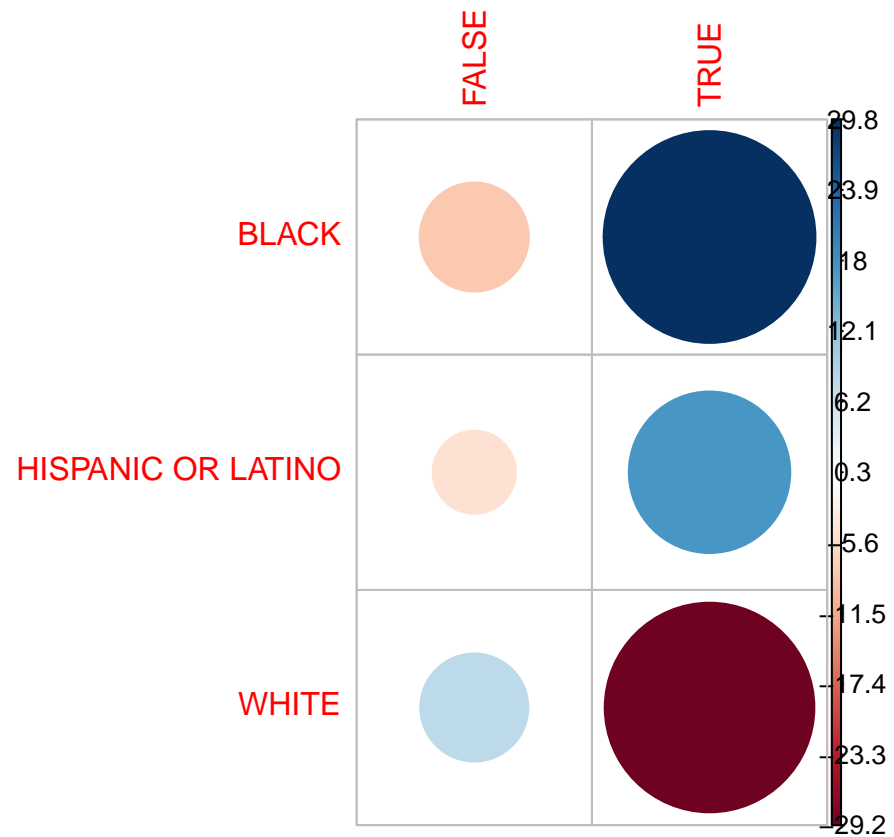|	FALSE	TRUE
FALSE	90360	6401
TRUE	788	24

```
knitr::kable(table(moving_traffic_violations$Race,moving_traffic_violations$Searched))
```

|	FALSE	TRUE
BLACK	12538	1867
HISPANIC OR LATINO	28535	2856
WHITE	50118	1708

```
race_searched_chisq <- chisq.test(table(moving_traffic_violations$Race,moving_traffic_violations$Searche
```

```r
corrplot(race_searched_chisq$residuals, is.cor = FALSE)
```



```r
moving_traffic_violations$RACE_searched <- paste0(moving_traffic_violations$Race," (",moving_traffic_vio

race_searched_type_tbl <- as.data.frame.matrix(table(moving_traffic_violations$RACE_searched,moving_tra

race_searched_type_tbl <- race_searched_type_tbl[c(1,3,4,6,7,9),]

corrplot(chisq.test(race_searched_type_tbl)$residuals, is.cor = FALSE)
```
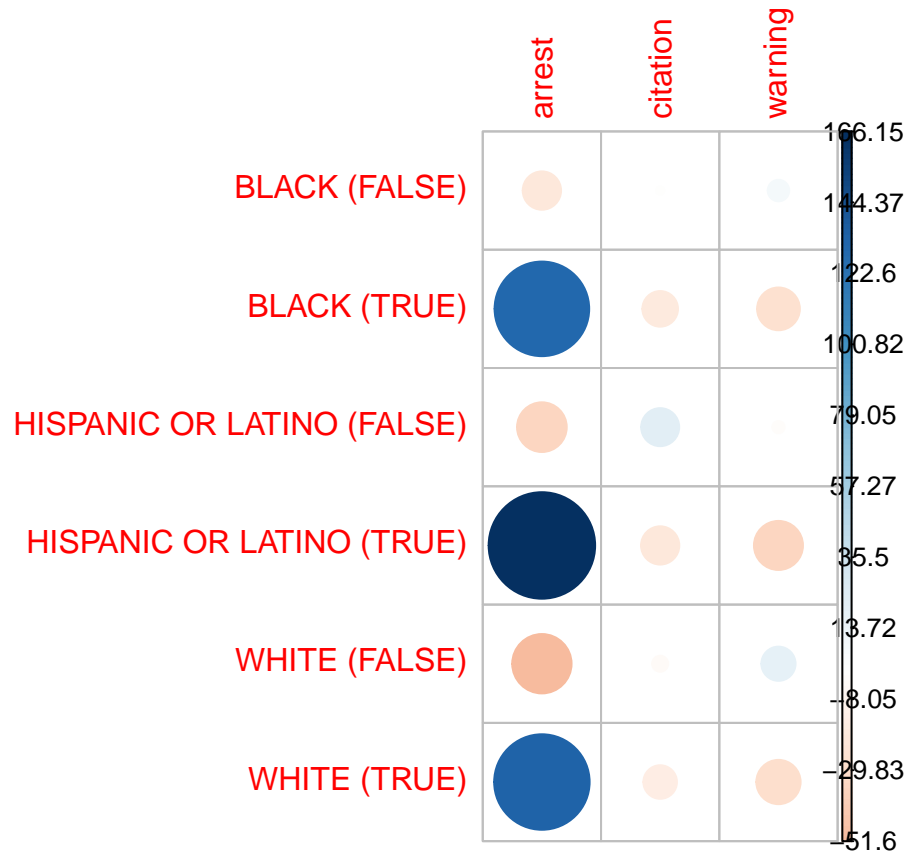
```
knitr::kable(table(moving_traffic_violations$RACE_searched,moving_traffic_violations$type))
```

	arrest	citation	warning
BLACK (FALSE)	147	3040	9351
BLACK (NA)	188	0	0
BLACK (TRUE)	1470	62	335
HISPANIC OR LATINO (FALSE)	183	8807	19545
HISPANIC OR LATINO (NA)	313	0	0
HISPANIC OR LATINO (TRUE)	2301	136	419
WHITE (FALSE)	121	11939	38058
WHITE (NA)	178	0	0
WHITE (TRUE)	1426	83	199

```
knitr::kable(t(table(moving_traffic_violations$type,moving_traffic_violations$Search_based_on)))
```

	arrest	citation	warning
	2604	23786	66954
ARREST OF PERSON IN VEHICLE	0	16	0
CONSENT	2	3	15
CONTRABAND/EVIDENCE IN PLAIN VIEW	22	13	7
FRISK FOR SAFETY	127	97	565
INCIDENTAL TO ARREST	2305	0	7
INVENTORY OF TOWED VEHICLE	2	0	0
PROBABLE CAUSE	1265	92	359

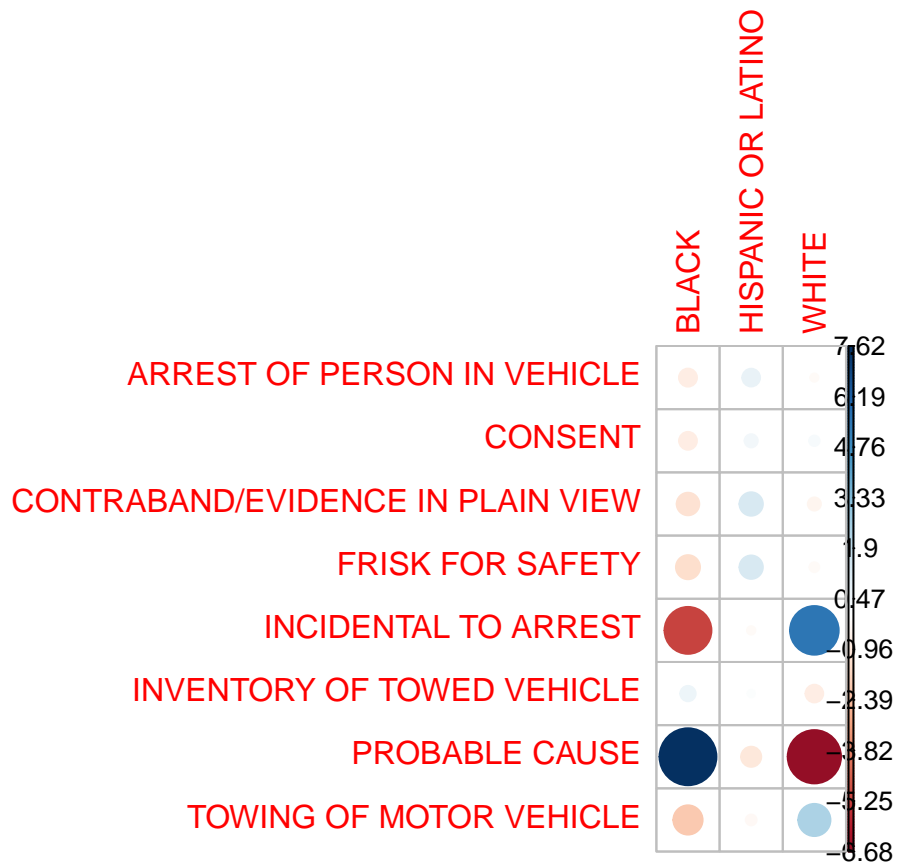|	arrest	citation	warning
TOWING OF MOTOR VEHICLE	0	60	0

```r
knitr::kable(t(table(moving_traffic_violations$Race,moving_traffic_violations$Search_based_on)))
```

	BLACK	HISPANIC OR LATINO	WHITE
	13157	29523	50664
ARREST OF PERSON IN VEHICLE	3	9	4
CONSENT	4	10	6
CONTRABAND/EVIDENCE IN PLAIN VIEW	8	24	10
FRISK FOR SAFETY	208	371	210
INCIDENTAL TO ARREST	536	1012	764
INVENTORY OF TOWED VEHICLE	1	1	0
PROBABLE CAUSE	667	729	320
TOWING OF MOTOR VEHICLE	9	25	26

```r
out <- chisq.test(t(as.data.frame.matrix(table(moving_traffic_violations$Race,moving_traffic_violations$
```

```
## Warning in
## chisq.test(t(as.data.frame.matrix(table(moving_traffic_violations$Race, : Chi-
## squared approximation may be incorrect
```

```r
corrplot(out$residuals, is.cor = FALSE)
```
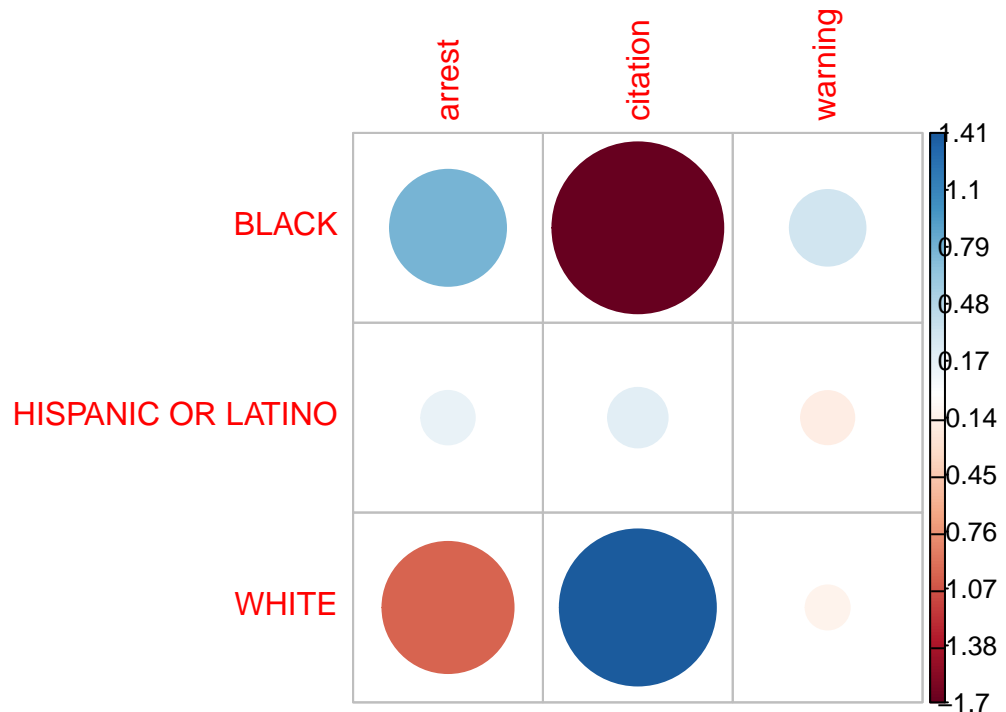
```
moving_traffic_violations_frisked <- moving_traffic_violations[which(moving_traffic_violations$Search_ba
```

```
knitr::kable(table(moving_traffic_violations_frisked$Race, moving_traffic_violations_frisked$type))
```

	arrest	citation	warning
BLACK	38	17	153
HISPANIC OR LATINO	61	47	263
WHITE	28	33	149

```
moving_traffic_violations_frisked.chisq <- chisq.test(table(moving_traffic_violations_frisked$Race, mov
```

```
moving_traffic_violations_frisked.chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(moving_traffic_violations_frisked$Race, moving_traffic_violations_frisked$type)
## X-squared = 6.6971, df = 4, p-value = 0.1528
```

```
corrplot(moving_traffic_violations_frisked.chisq$residuals, is.cor = FALSE)
```



```
# Create seperate searched dataset
```

```
frisk_found_tbl <- as.data.frame.matrix(table(moving_traffic_violations_frisked$Race,moving_traffic_viol
```

```
knitr::kable(frisk_found_tbl)
```

	CASH	DRUGS	NOTHING	OTHER	WEAPONS
BLACK	2	1	172	16	17
HISPANIC OR LATINO	1	1	314	38	17
WHITE	0	1	175	14	19

```
chisq.test(frisk_found_tbl)
```

```
## Warning in chisq.test(frisk_found_tbl): Chi-squared approximation may be
## incorrect
##
##  Pearson's Chi-squared test
##
## data:  frisk_found_tbl
## X-squared = 10.202, df = 8, p-value = 0.2512
```

```
black_found_prop <- c(frisk_found_tbl[1,3], sum(frisk_found_tbl[1,]))
hispanic_found_prop <- c(frisk_found_tbl[2,3], sum(frisk_found_tbl[2,]))
white_found_prop <- c(frisk_found_tbl[3,3], sum(frisk_found_tbl[3,]))
total_prop <- c(sum(frisk_found_tbl[,3]), sum(rowSums(frisk_found_tbl)))

prop.test(c(black_found_prop[1],total_prop[1]),c(black_found_prop[2],total_prop[2]))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(black_found_prop[1], total_prop[1]) out of c(black_found_prop[2], total_prop[2])
## X-squared = 0.094616, df = 1, p-value = 0.7584
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.07241341  0.04859459
## sample estimates:
##    prop 1    prop 2
## 0.8269231 0.8388325
```

```
prop.test(c(hispanic_found_prop[1],total_prop[1]),c(hispanic_found_prop[2],total_prop[2]))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(hispanic_found_prop[1], total_prop[1]) out of c(hispanic_found_prop[2], total_prop[2])
## X-squared = 0.058102, df = 1, p-value = 0.8095
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03923602  0.05429342
## sample estimates:
##    prop 1    prop 2
## 0.8463612 0.8388325
```

```
prop.test(c(white_found_prop[1],total_prop[1]),c(white_found_prop[2],total_prop[2]))
```

```
##
##  2-sample test for equality of proportions with continuity correction
```

```
## 
## data:  c(white_found_prop[1], total_prop[1]) out of c(white_found_prop[2], total_prop[2])
## X-squared = 6.3207e-30, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.05926181  0.05623798
## sample estimates:
##    prop 1    prop 2
## 0.8373206 0.8388325
```

```r
# Proportion tests for searchs
```