StatsProject

Bowen Jiang

Comparison between 2010 and 2015

1. Data Preprocessing

```
sqf2010=read.csv("2010_sqf_m35.csv")
sqf2015=read.csv("2015_sqf_m35.csv")

sqf2015=sqf2015[!sqf2015$perstop=="**" & !sqf2015$perstop==" ",]
sqf2015$perstop=as.numeric(as.character(sqf2015$perstop))
sqf2010$perstop = as.numeric(as.character(sqf2010$perstop))
```

2. Overview

We are mainly interested in how the composition of people arrested changes between 2010 and 2015, espeically in terms of races.

We can first look at some basic information of the two years.

The number of people who got arrested in 2010 is,

```
arst2010=sqf2010$arstmade
sum(arst2010)
```

```
## [1] 41084
```

Out of a population size of

```
length(arst2010)
```

```
## [1] 601285
```

Therefore, the percentage of arrest is around 6.83%

```
sum(arst2010)/length(arst2010)
```

```
## [1] 0.068327
```

Similarly, for 2015,

```
arst2015=sqf2015$arstmade
sum(arst2015)
```

```
## [1] 3936
```

```
length(arst2015)
```

```
## [1] 22502
```

```
sum(arst2015)/length(arst2015)
```

```
## [1] 0.1749178
```

We have 3936 people who got arrested out of 22502, the arrest rate is around 17.49%

3.

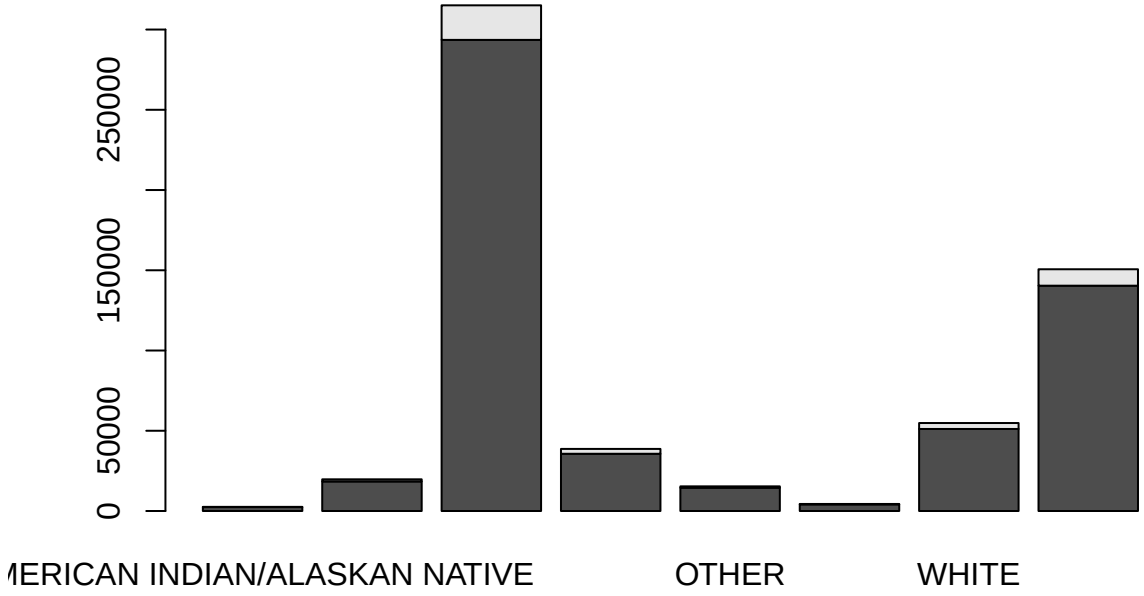
We can then look at the breakdown of arrested population in race.

In 2010, we have,

```
counts2010=table(sqf2010$arstmade,sqf2010$race)
counts2010
```

```
##
##     AMERICAN INDIAN/ALASKAN NATIVE ASIAN/PACIFIC ISLANDER  BLACK
##   0                           2453                  18411 293555
##   1                            126                   1321  21528
##
##     BLACK-HISPANIC  OTHER UNKNOWN  WHITE WHITE-HISPANIC
##   0          35657  14505    4114  51112          140394
##   1           3032    855     281   3698           10243
```

```
barplot(counts2010)
```



we can also convert the data into percentage

```
perc2010=prop.table(counts2010,2)
perc2010
```

```
##
##     AMERICAN INDIAN/ALASKAN NATIVE ASIAN/PACIFIC ISLANDER      BLACK
##   0                     0.95114385              0.93305291 0.93167515
##   1                     0.04885615              0.06694709 0.06832485
##
##     BLACK-HISPANIC      OTHER    UNKNOWN      WHITE WHITE-HISPANIC
##   0     0.92163147 0.94433594 0.93606371 0.93253056     0.93200210
##   1     0.07836853 0.05566406 0.06393629 0.06746944     0.06799790
```

then we have a percentage of arrested for each race,

American Indian/Alaskan Native: 4.89%

Asian/Parcific Islander: 6.69%

Black: 6.83%

Black-hispanic: 7.83%

Other: 5.57%

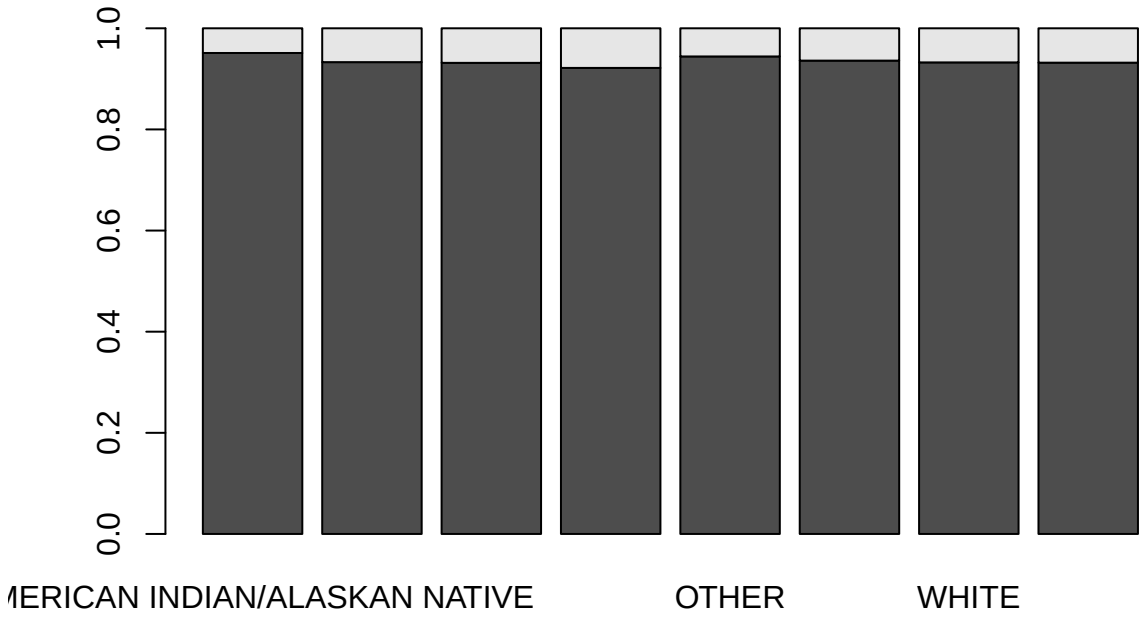
Unknown: 6.39%

White: 6.75%

White-hispanic: 6.80%

which we can also plot on a barplot,

```
barplot(perc2010)
```

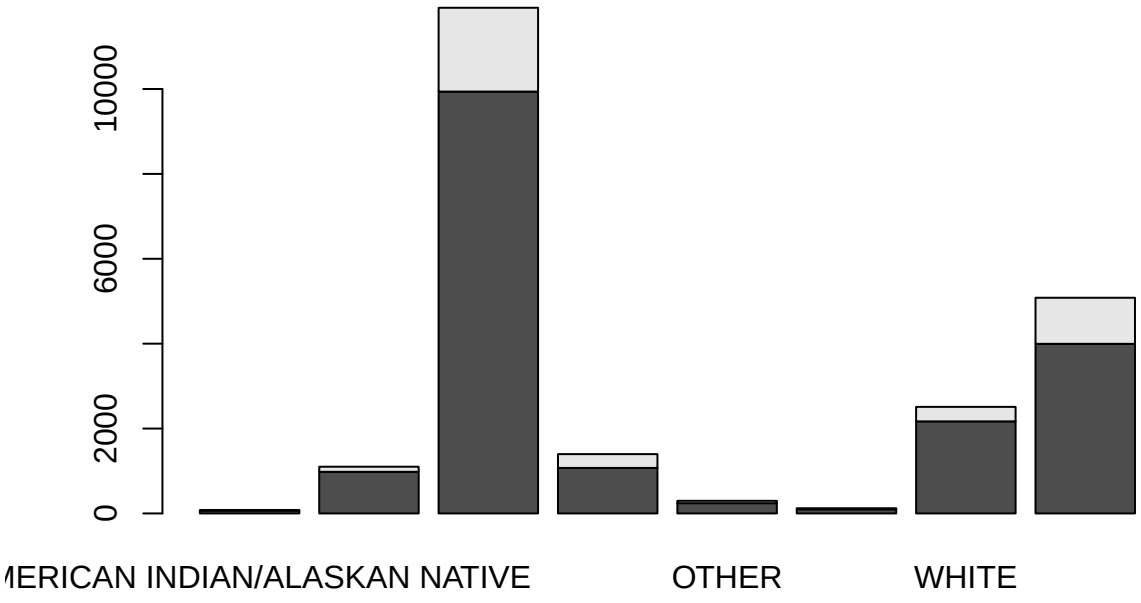


similarly, we calculate the statistics of 2015,

```
counts2015=table(sqf2015$arstmade,sqf2015$race)
counts2015
```

```
##
##     AMERICAN INDIAN/ALASKAN NATIVE ASIAN/PACIFIC ISLANDER BLACK
##   0                             69                    981  9939
##   1                              8                    120  1975
##
##     BLACK-HISPANIC OTHER UNKNOWN WHITE WHITE-HISPANIC
##   0           1073   242      97  2169           3996
##   1            325    55      25   342           1086
```

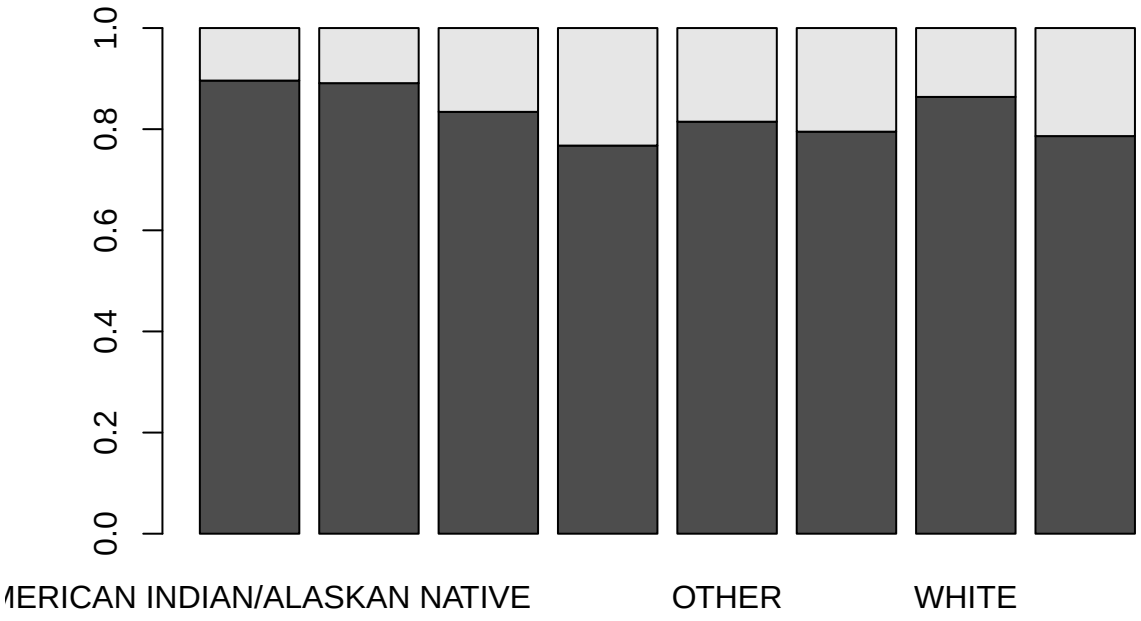
```
barplot(counts2015)
```




```
perc2015=prop.table(counts2015,2)
perc2015
```

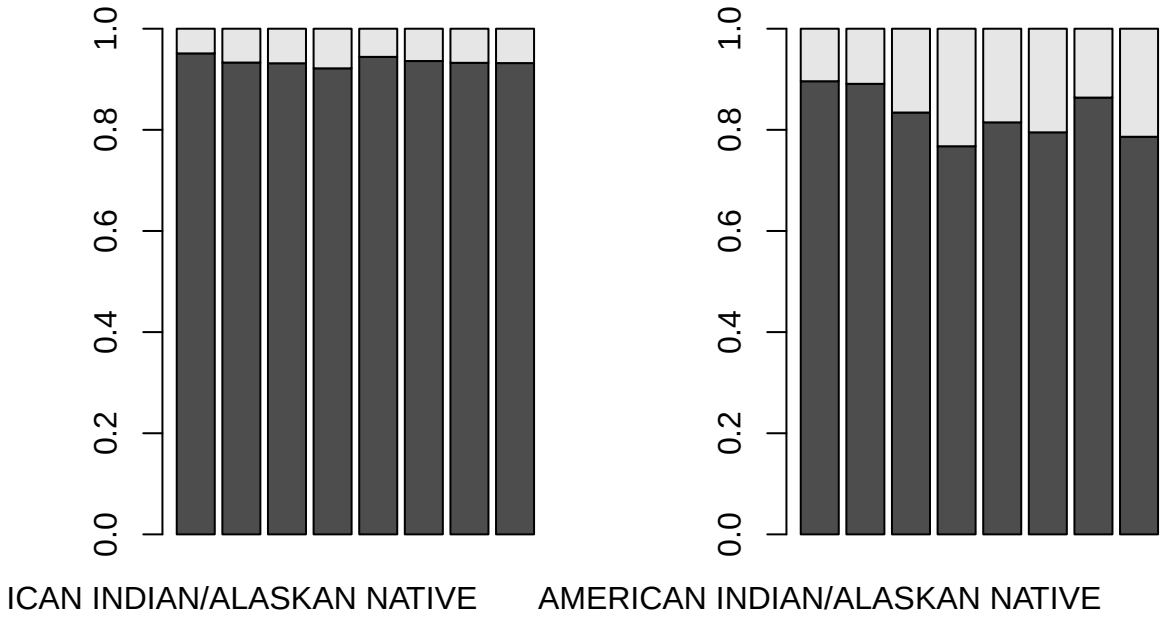
```
##
##     AMERICAN INDIAN/ALASKAN NATIVE ASIAN/PACIFIC ISLANDER     BLACK
##   0                      0.8961039              0.8910082 0.8342286
##   1                      0.1038961              0.1089918 0.1657714
##
##     BLACK-HISPANIC     OTHER   UNKNOWN     WHITE WHITE-HISPANIC
##   0      0.7675250 0.8148148 0.7950820 0.8637993      0.7863046
##   1      0.2324750 0.1851852 0.2049180 0.1362007      0.2136954
```

```
barplot(perc2015)
```



To visualize the difference, we put the two bar plots of percentage together,

```
par(mfrow=c(1,2))
barplot(perc2010)
barplot(perc2015)
```



ICAN INDIAN/ALASKAN NATIVE AMERICAN INDIAN/ALASKAN NATIVE

```
par(mfrow=c(1,2))
```

For year 2010, the race that is most likely to be arrested is black-hispanic, and it stays the same for year 2015.