# US Homicides Ethnical Analysis

*Rodrigo Scroferneker*

*13/12/2017*

## Introduction

This three step analysis academic work has the purpose of investigating the homicide statistics in the United States. Using a FBI database that goes from 1980-2014 that contains the number of homicides and details according to each case such as state where it was committed, ethnicity of the murderer, ethnicity of the victim, race, the level of relationship of the involved, age, sex and murder weapon, this work aims to answer two questions focusing on a racial approach: is there a difference between the number of solved cases when it comes to race? What state kills more by race? Therefore, this work will do a more social inclined take on the dataset.

## Objective

This assignment will start by only sorting out the amount of solved and unsolved cases in general for the whole country. Then it will continue by analyzing the official statistics with the main intention of finding out if there are any divergences on the amount of solved and unsolved cases depending on the race of the victim, focusing on four major races: White, Asian/Pacific islander, native American/Alaska native and Black. To close if of, there will be a more specific approach to each one of the 50 states to identify which of the races has a higher murder statistic.

## Homicides in General

```
df0 <- df %>%
  group_by(Crime.Solved) %>%
  summarise(N = n()) %>%
  summarise(Total = sum(N), NotSolved = N[1], Solved = N[2], Perc.Not.Solved = NotSolved*100/Total)
grid.table(df0)
```

| | Total | NotSolved | Solved | Perc.Not.Solved |
|---|---|---|---|---|
| *1* | 638454 | 190282 | 448172 | 29.8035567166938 |

As it can be noted, our sample has a great ammount of data for each one of the states from 1980-2014. Here it shows that 29.80% of the cases are left unsolved.

```
grid.table(df4)
```

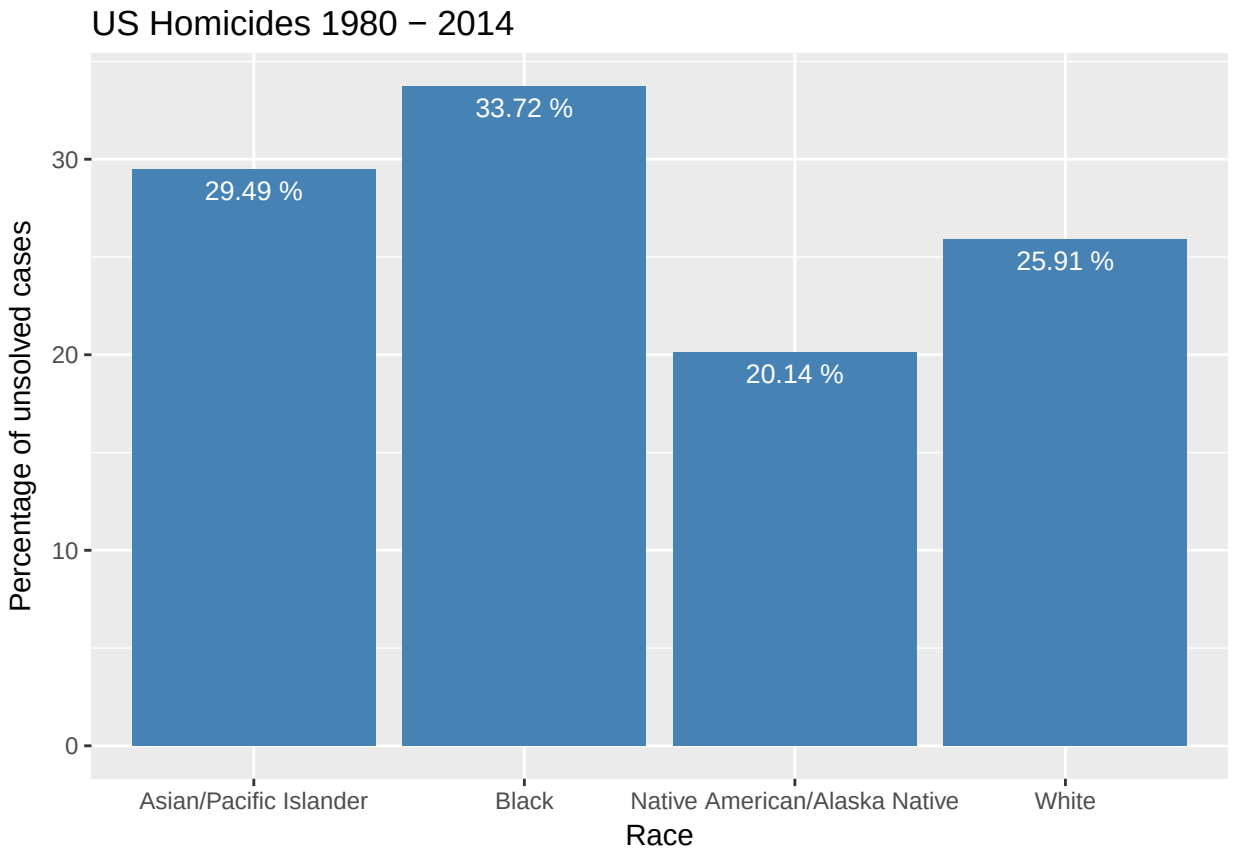
| | Victim.Race | Not.Solved | Solved | Total | Perc.Not.Solved |
|---|---|---|---|---|---|
| *1* | Black | 101125 | 198774 | 299899 | 33.71969 |
| *2* | Asian/Pacific Islander | 2917 | 6973 | 9890 | 29.49444 |
| *3* | White | 82236 | 235186 | 317422 | 25.90747 |
| *4* | Native American/Alaska Native | 920 | 3647 | 4567 | 20.14451 |

Looking into this graphic, we can see that there is some sort of difference between the amount of unsolved cases between black and white people even though they have nearly the same amount of cases. When it comes to the Asian/Pacific Islander e Native American/Alaska Native, they have a smaller rate, but they also have a way less significant amount of cases, so it isn't that simple to make assertives about those two races when compared to the white and black rate of unsolved crimes.

```
library(ggplot2)

specify_decimal <- function(x, k) trimws(format(round(x, k), nsmall=k))

df4 %>%ggplot(aes(x = Victim.Race, y = Perc.Not.Solved)) +
  geom_bar(stat="identity", fill="steelblue") +
  geom_text(label = paste(specify_decimal(df4$Perc.Not.Solved, 2),"%"),  vjust=1.6, color="white", size=
  labs(y = "Percentage of unsolved cases", x = "Race", title = "US Homicides 1980 - 2014")
```




## Unsolved cases by state

We are looking into making a more specific approch to the dataset of solved and unsolved murders so, from now on, each percentage will be sorted out by state.

```
  df5 <- df %>% filter(Victim.Race != "Unknown") %>%
    group_by(State, Crime.Solved) %>%
    summarise( N = n()) %>%
    mutate(region=tolower(State))

notSolved <-data.frame(filter(df5, Crime.Solved == "No") %>%
```

```
    mutate(Not.Solved = N)   %>%
    select("Not.Solved", "region"))
```

```
## Adding missing grouping variables: `State`
```

```
solved <- data.frame (filter(df5, Crime.Solved == "Yes") %>%
    mutate(Solved = N)   %>%
    select("Solved", "region"))
```

```
## Adding missing grouping variables: `State`
```

```
df5 <- merge(x = notSolved, y = solved, by = c("region"))
df5 <- df5 %>% mutate(Total =  df5$Not.Solved + Solved, Perc.Not.Solved = df5$Not.Solved * 100/ Total)

df5 %>% head(5) %>% grid.table()
```

| | Perc.Not.Solved | region |
|---|---|---|
| 1 | 65.75381 | district of columbia |
| 2 | 45.81784 | new york |
| 3 | 40.37436 | maryland |
| 4 | 38.72191 | illinois |
| 5 | 36.74750 | massachusetts |

As we can see, here is the top 5 states with the biggest amount of unsolved crimes.

## Unsolved cases by race

```
df5 <- df %>%
  group_by(State, Crime.Solved, Victim.Race) %>%
  summarise( N = n()) %>%
  mutate(region=tolower(State))

notSolved <-data.frame (filter(df5, Crime.Solved == "No") %>%
  mutate(Not.Solved = N)   %>%
  filter( Victim.Race != "Unknown")) %>%
  select("Victim.Race", "Not.Solved", "region")

solved <- data.frame (filter(df5, Crime.Solved == "Yes") %>%
  mutate(Solved = N)   %>%
  filter( Victim.Race != "Unknown")) %>%
  select("Victim.Race", "Solved", "region")

df5 <- merge(x = notSolved, y = solved, by = c("region", "Victim.Race"))
df5 <- df5 %>% mutate(Total =  df5$Not.Solved + Solved, Perc.Not.Solved = df5$Not.Solved * 100/ Total)

lapply(races, function(str){
  df5 %>% filter(Victim.Race == str) %>% arrange(desc(Perc.Not.Solved)) %>% select("region","Victim.Rac
})
```

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

```
##                  region Victim.Race Perc.Not.Solved
## 1  district of columbia       White        62.42545
## 2              new york       White        42.99229
## 3              illinois       White        35.80310
## 4            california       White        34.19786
## 5         rhodes island       White        29.90431
## 6                nevada       White        29.37970
## 7            new jersey       White        28.87496
## 8                arizona       White        27.62555
## 9           connecticut       White        27.19569
## 10        massachusetts       White        27.00333
##
## [[2]]
##                  region Victim.Race Perc.Not.Solved
## 1  district of columbia       Black        65.97212
## 2         massachusetts       Black        50.74257
## 3              new york       Black        48.31065
## 4              maryland       Black        45.43508
## 5              missouri       Black        44.63073
## 6            california       Black        41.15258
## 7           connecticut       Black        40.90699
## 8            new jersey       Black        40.77162
## 9              illinois       Black        40.10232
## 10        rhodes island       Black        39.57055
##
## [[3]]
##                  region            Victim.Race Perc.Not.Solved
## 1  district of columbia Asian/Pacific Islander        70.21277
## 2              new york Asian/Pacific Islander        42.12766
## 3             louisiana Asian/Pacific Islander        37.12121
## 4              maryland Asian/Pacific Islander        35.84906
## 5            california Asian/Pacific Islander        34.34658
## 6         new hampshire Asian/Pacific Islander        33.33333
## 7         west virginia Asian/Pacific Islander        33.33333
## 8               georgia Asian/Pacific Islander        32.25806
## 9              michigan Asian/Pacific Islander        31.76471
## 10             illinois Asian/Pacific Islander        31.44654
##
## [[4]]
##                  region                    Victim.Race Perc.Not.Solved
## 1  district of columbia Native American/Alaska Native        71.42857
## 2           connecticut Native American/Alaska Native        50.00000
## 3              new york Native American/Alaska Native        50.00000
## 4          pennsylvania Native American/Alaska Native        41.17647
## 5                hawaii Native American/Alaska Native        33.33333
## 6           mississippi Native American/Alaska Native        33.33333
## 7                 texas Native American/Alaska Native        32.55814
## 8                  ohio Native American/Alaska Native        30.76923
## 9            new jersey Native American/Alaska Native        29.26829
## 10           new mexico Native American/Alaska Native        29.15361
```

Here we can see the top 10 of unsolved cases divided by race.

```
library(ggplot2)
library(maps)
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
us <- map_data("state")

f <- function(str) {

  df6 <- df5 %>% filter(Victim.Race == str)
  gg <- ggplot()

  gg <- gg + geom_map(data=us, map=us,
                      aes(x=long, y=lat, map_id=region),
                      fill="#ffffff", color="#ffffff", size=0.15)

  gg <- gg + geom_map(data=df6, map=us,
                      aes(fill=Perc.Not.Solved, map_id=region),
                      color="#ffffff", size=0.15)

  gg <- gg + scale_fill_gradient(low="white", high="darkblue", name="")

  gg <- gg + labs(x="", y="", title = paste("Rate (in %) of unsolved cases by state:", str)) +
      coord_map("albers", lat0 = 39, lat1 = 45) +
      theme(panel.border = element_blank()) +
      theme(panel.background = element_blank()) +
      theme(axis.ticks = element_blank()) +
      theme(axis.text = element_blank())
  gg
}

lapply(races, f)
```

```
## Warning: Ignoring unknown aesthetics: x, y
```

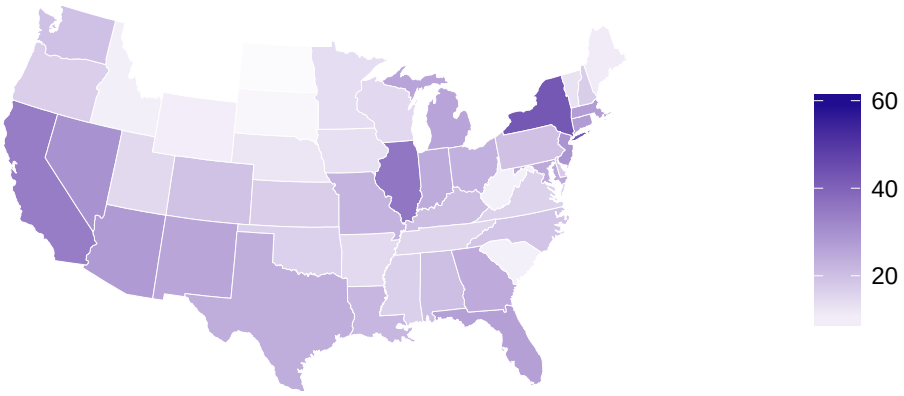
```
## Warning: Ignoring unknown aesthetics: x, y
```

```
## Warning: Ignoring unknown aesthetics: x, y
```

```
## Warning: Ignoring unknown aesthetics: x, y
```

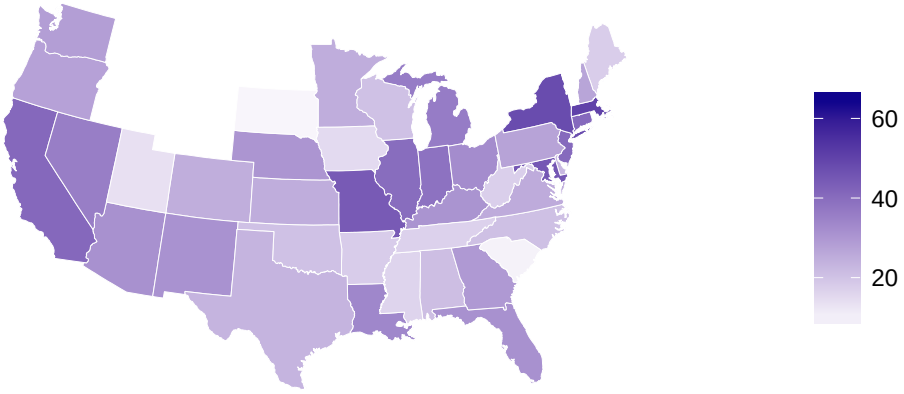
```
## [[1]]
```

Rate (in %) of unsolved cases by state: White



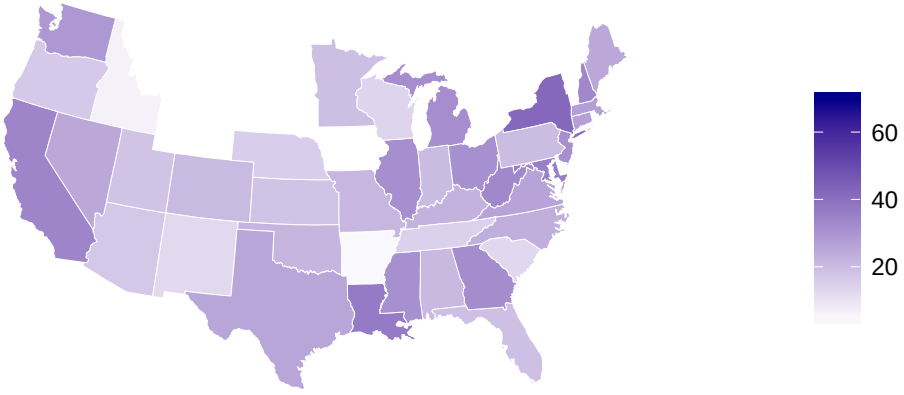

```
##
## [[2]]
```

Rate (in %) of unsolved cases by state: Black



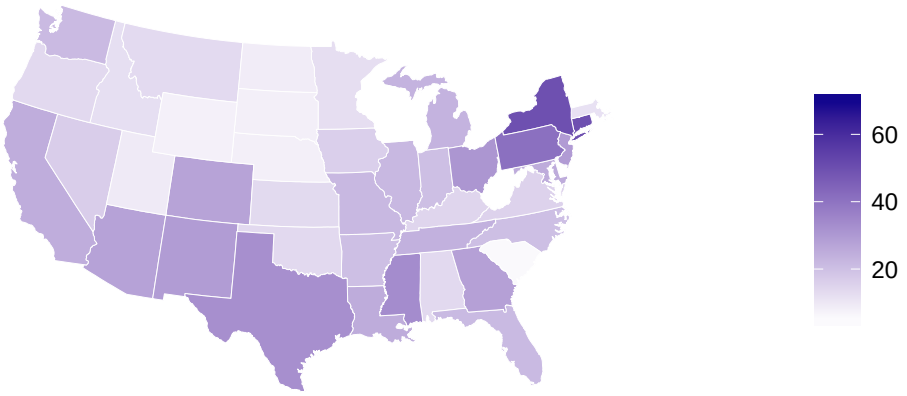

```
##
## [[3]]
```

Rate (in %) of unsolved cases by state: Asian/Pacific Islander




```
##
## [[4]]
```

Rate (in %) of unsolved cases by state: Native American/Alaska Native




We can observe that, even with differerents intensities of color in each state, the statates follows some pattern according to the race.

## Conclusion

After a final dissection of the dataset, its possible to conclude that the numbers of solved and unsolved cases for each race in each state are in a way - except some places - proportional, meaning that, when looking into the efficiency of the system to solve cases they are either very efficient or not efficient despite the race. Take the District of Columbia for example, when it comes to crimes of manslaughter committed against black people 65.97% of the cases are unsolved, a more alarming rate of 71.42% when it comes to native Americans, 70.21% when it comes to Asian and pacific islanders and a rate of 62.42% when it comes to white people, the average of unsolved cases in the District of Columbia hits a total of 65.75% , more than half of the murders committed on that state are left unsolved and that is an issue that affects every race in very high rates. It's important to also point out the startling and, when it comes to the United States of America in general, out of the pattern discrepancy that happens is Massachusetts. Massachusetts is the fifth state with the largest amount of unsolved cases in the whole country, 36.74% of its murders are left without an answer, but a deeper analysis of the race statists will show that 50.74% of the cases regarding black people are unsolved, ranking as second, however, it's only the tenth state with the largest amount of unsolved cases regarding white people (27% of the cases) and it doesn't even show up on the top ten for native Americans/Alaska natives and Asian/Pacific islander rates. On account of the data analysis and the statistics presented in this work, we can see that the differences in the number of solved and unsolved cases for each race, in most of the country are very small. In some of the states, it is possible to see that there is a variance, but it's something that needs to be look into deeper, needing a research that approaches historical, social and other issues that are not presented on the FBI dataset to find out why those variances can be noted.