

HOMICIDE REPORTS

TABLE OF CONTENTS



- | | | | |
|---|------------------------|---|------------------------------|
| 1 | DATASET
DESCRIPTION | 2 | EXPLORATORY
DATA ANALYSIS |
| 3 | INSIGHTS | 4 | PREDICTION |



THE DATASET: MURDER ACCOUNTABILITY PROJECT

- THE MOST COMPLETE DATABASE ON HOMICIDES IN THE US
- HOMICIDES COMMITTED BETWEEN 1976 AND 2014
- DATA WAS COLLECTED FROM THE FBI'S SUPPLEMENTARY HOMICIDE REPORT AND FREEDOM OF INFORMATION ACT



THE DATASET: MURDER ACCOUNTABILITY PROJECT

Column	Non-Null Count	Dtype
Record ID	638454	non-null int64
Agency Code	638454	non-null object
Agency Name	638454	non-null object
Agency Type	638454	non-null object
City	638454	non-null object
State	638454	non-null object
Year	638454	non-null int64
Month	638454	non-null object
Incident	638454	non-null int64
Crime Type	638454	non-null object
Crime Solved	638454	non-null object
Victim Sex	638454	non-null object
Victim Age	638454	non-null int64
Victim Race	638454	non-null object
Victim Ethnicity	638454	non-null object
Perpetrator Sex	638454	non-null object
Perpetrator Age	638454	non-null object
Perpetrator Race	638454	non-null object
Perpetrator Ethnicity	638454	non-null object
Relationship	638454	non-null object
Weapon	638454	non-null object
Victim Count	638454	non-null int64
Perpetrator Count	638454	non-null int64
Record Source	638454	non-null object

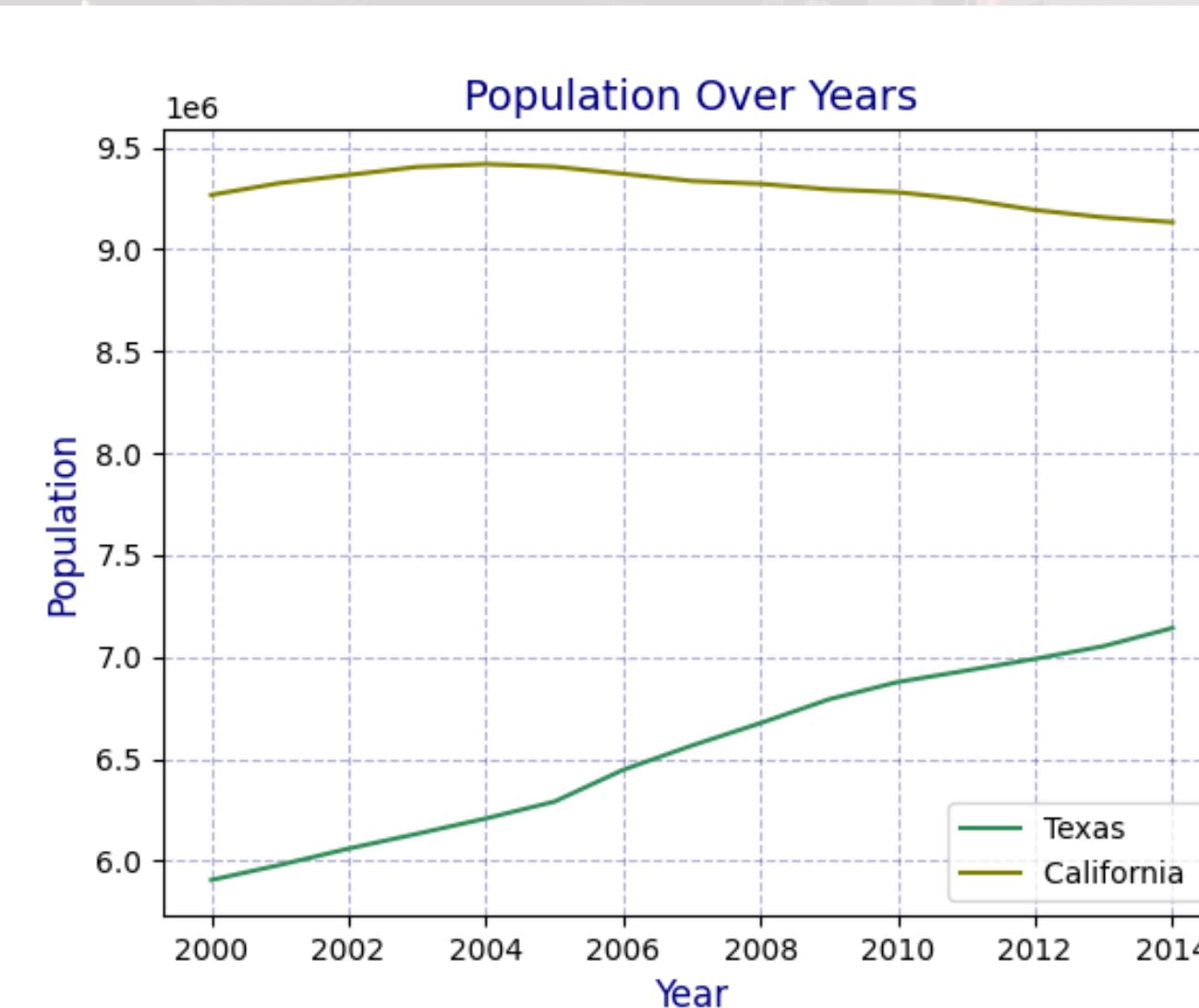


DATA CLEANING

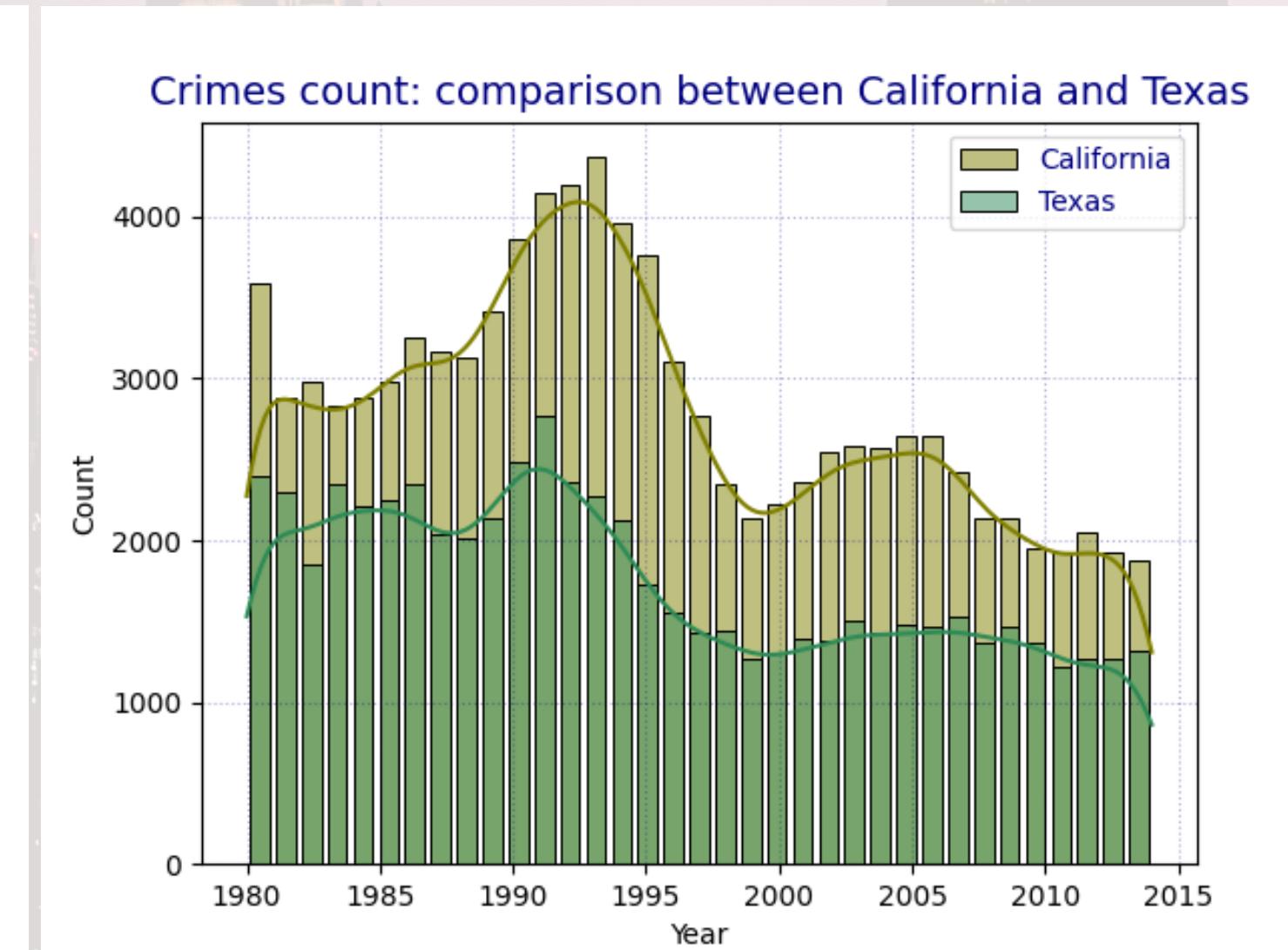
	count	mean	std	min	25%	50%	75%	max
Record ID	638454.0	319227.500000	184305.938720	1.0	159614.25	319227.5	478840.75	638454.0
Year	638454.0	1995.801102	9.927693	1980.0	1987.00	1995.0	2004.00	2014.0
Incident	638454.0	22.967924	92.149821	0.0	1.00	2.0	10.00	999.0
Victim Age	638454.0	35.033512	41.628306	0.0	22.00	30.0	42.00	998.0
Perpetrator Age	638453.0	20.322697	17.886842	0.0	0.00	21.0	31.00	99.0
Victim Count	638454.0	0.123334	0.537733	0.0	0.00	0.0	0.00	10.0
Perpetrator Count	638454.0	0.185224	0.585496	0.0	0.00	0.0	0.00	10.0

- NO MISSING VALUES
- A LOT OF UNKNOWN VALUES TO BE FILTERED OUT
- MANY OUTLIERS TO REMOVE - EX. MAX VICTIM AGE 998

HOMICIDES OVER THE YEARS



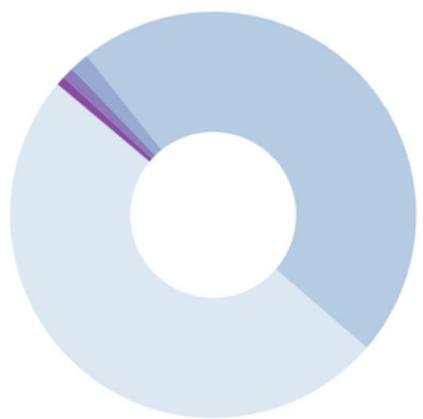
additional dataset



coherence with crises

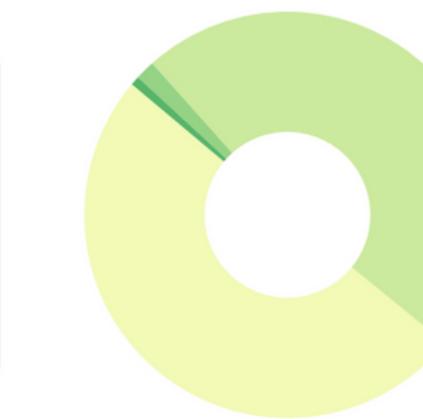
VICTIMS' RACES

Considering unknown race

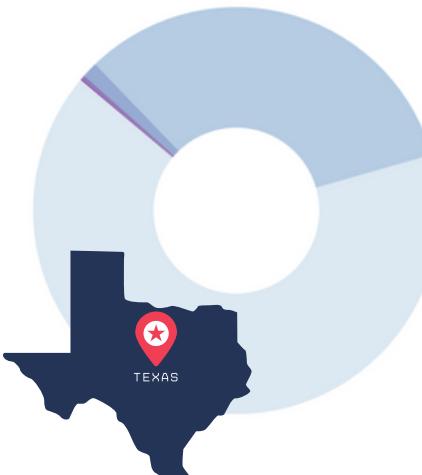


Legend	
White:	317285 (49.8%)
Black:	299788 (47.0%)
Asian/Pacific Islander:	9888 (1.6%)
Unknown:	5954 (0.9%)
Native American/Alaska Native:	4565 (0.7%)

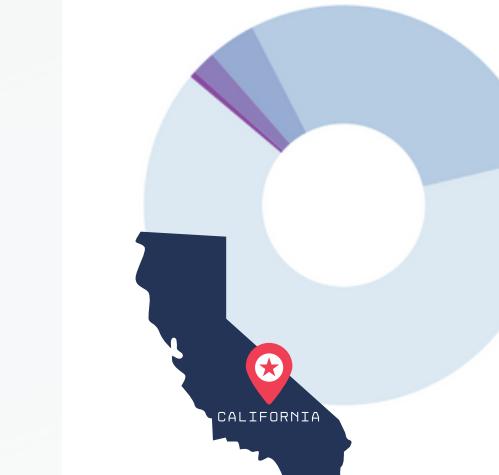
Dropping unknown race



Legend	
White:	317285 (50.2%)
Black:	299788 (47.5%)
Asian/Pacific Islander:	9888 (1.6%)
Native American/Alaska Native:	4565 (0.7%)

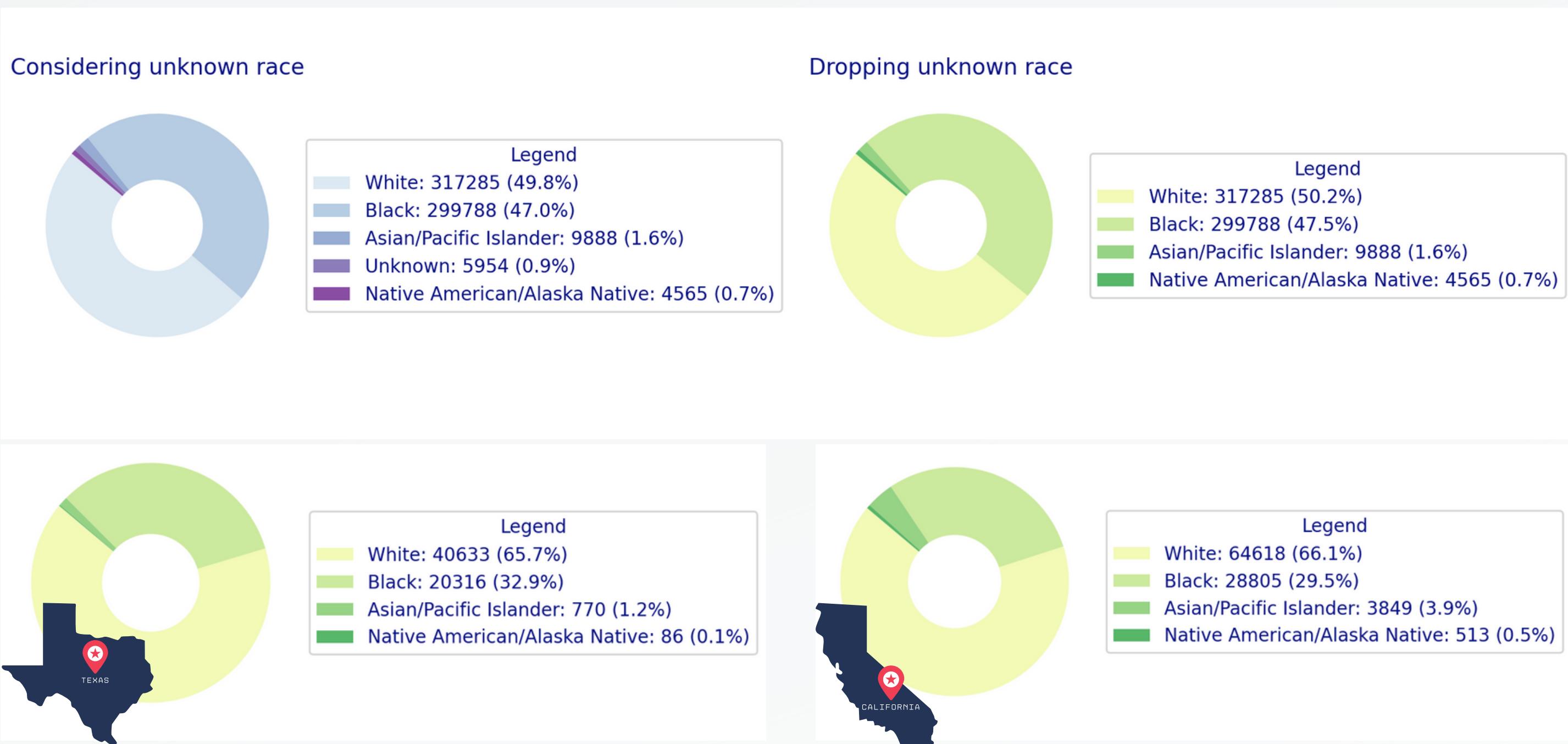


Legend	
White:	40633 (65.5%)
Black:	20316 (32.8%)
Asian/Pacific Islander:	770 (1.2%)
Unknown:	190 (0.3%)
Native American/Alaska Native:	86 (0.1%)

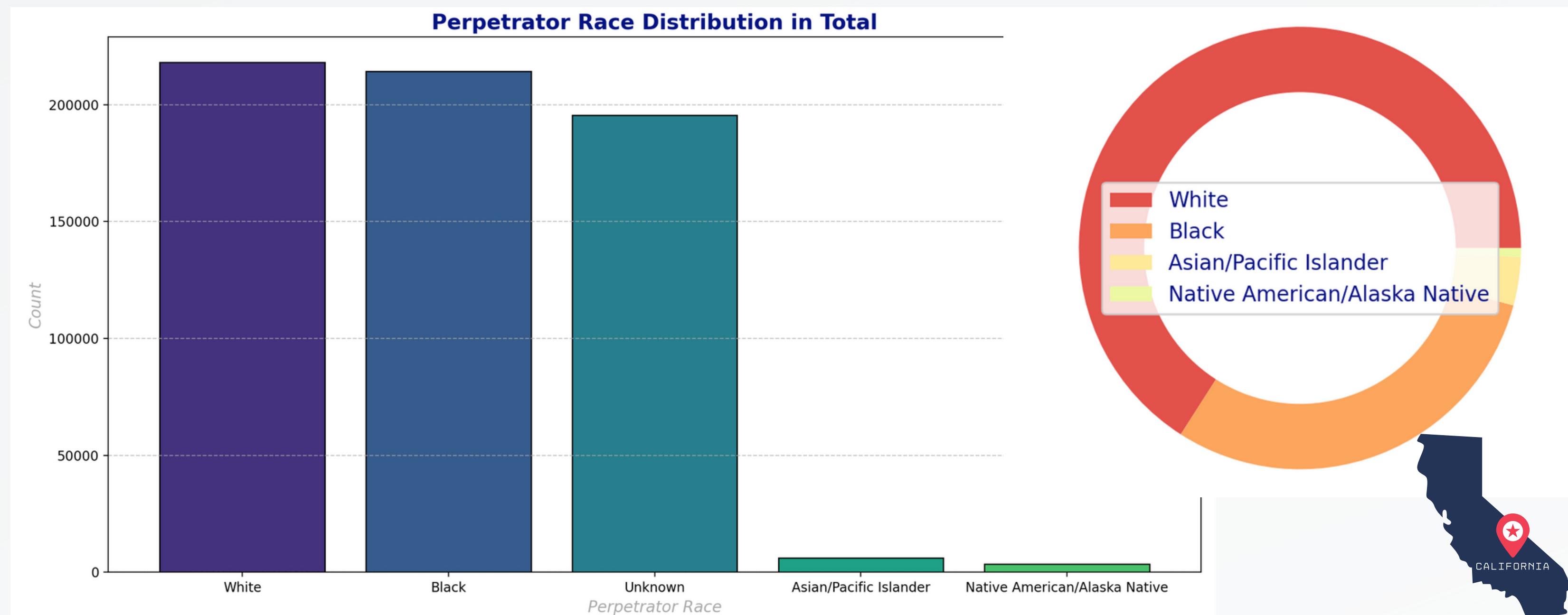


Legend	
White:	64618 (64.8%)
Black:	28805 (28.9%)
Asian/Pacific Islander:	3849 (3.9%)
Unknown:	1954 (2.0%)
Native American/Alaska Native:	513 (0.5%)

VICTIMS' RACES

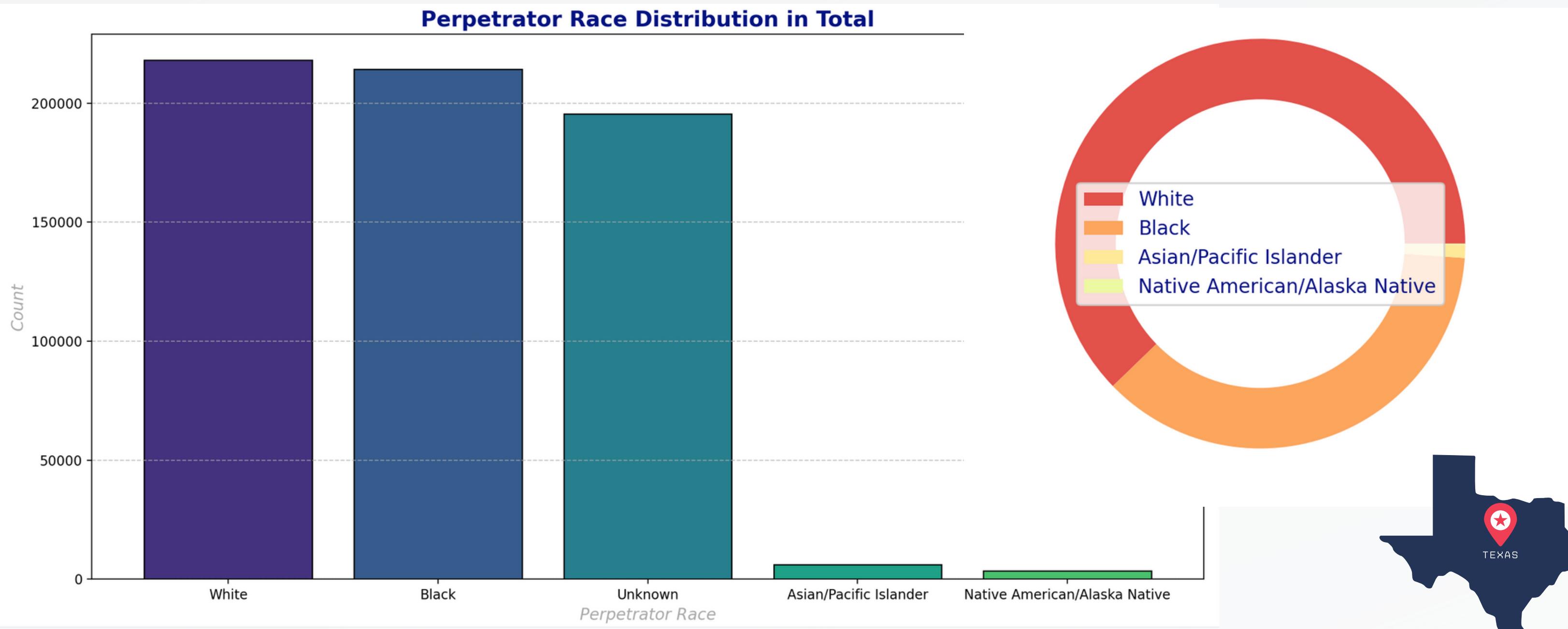


PERPETRATORS' RACES

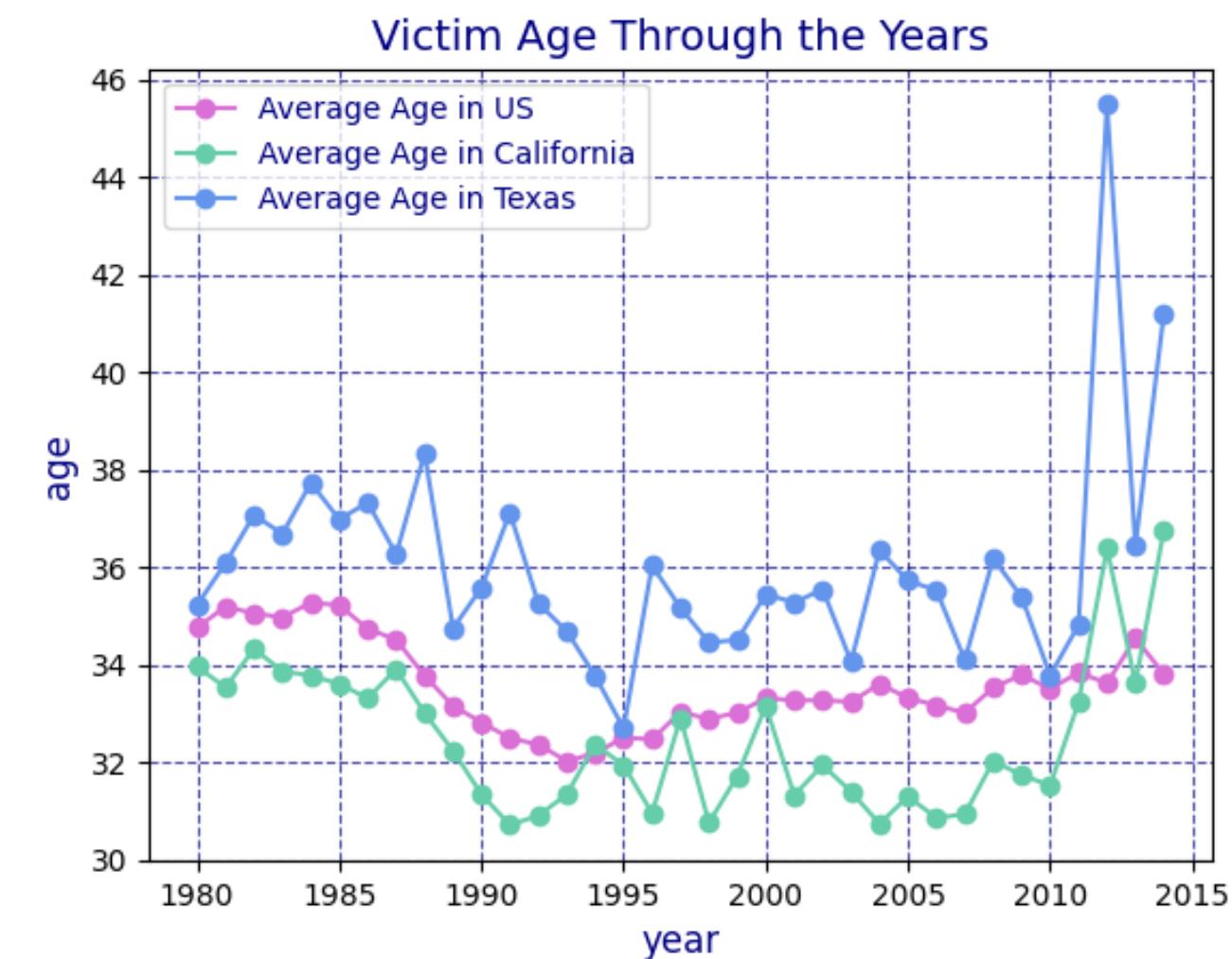
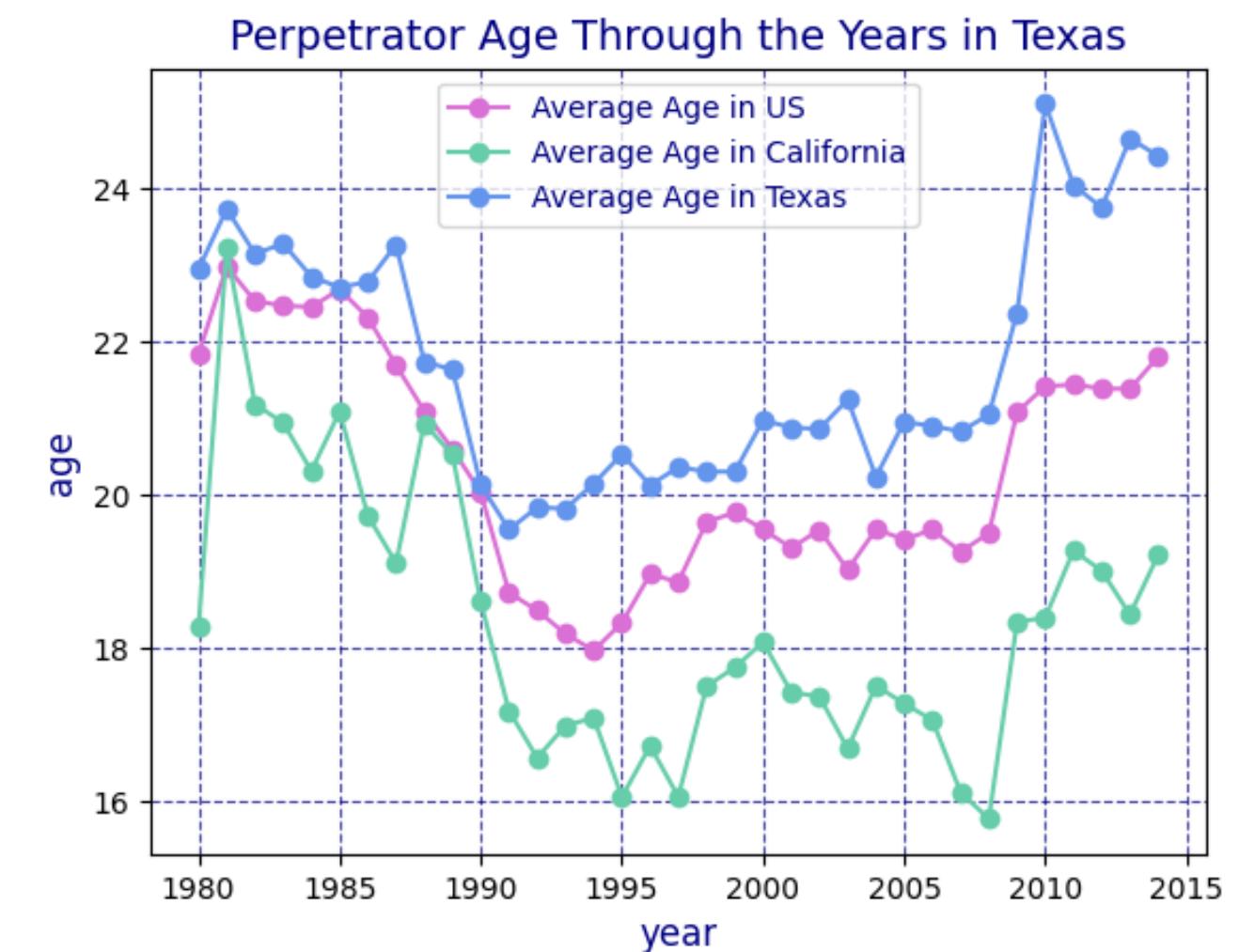


PERPETRATORS' RACES

Perpetrator Race Distribution in Total

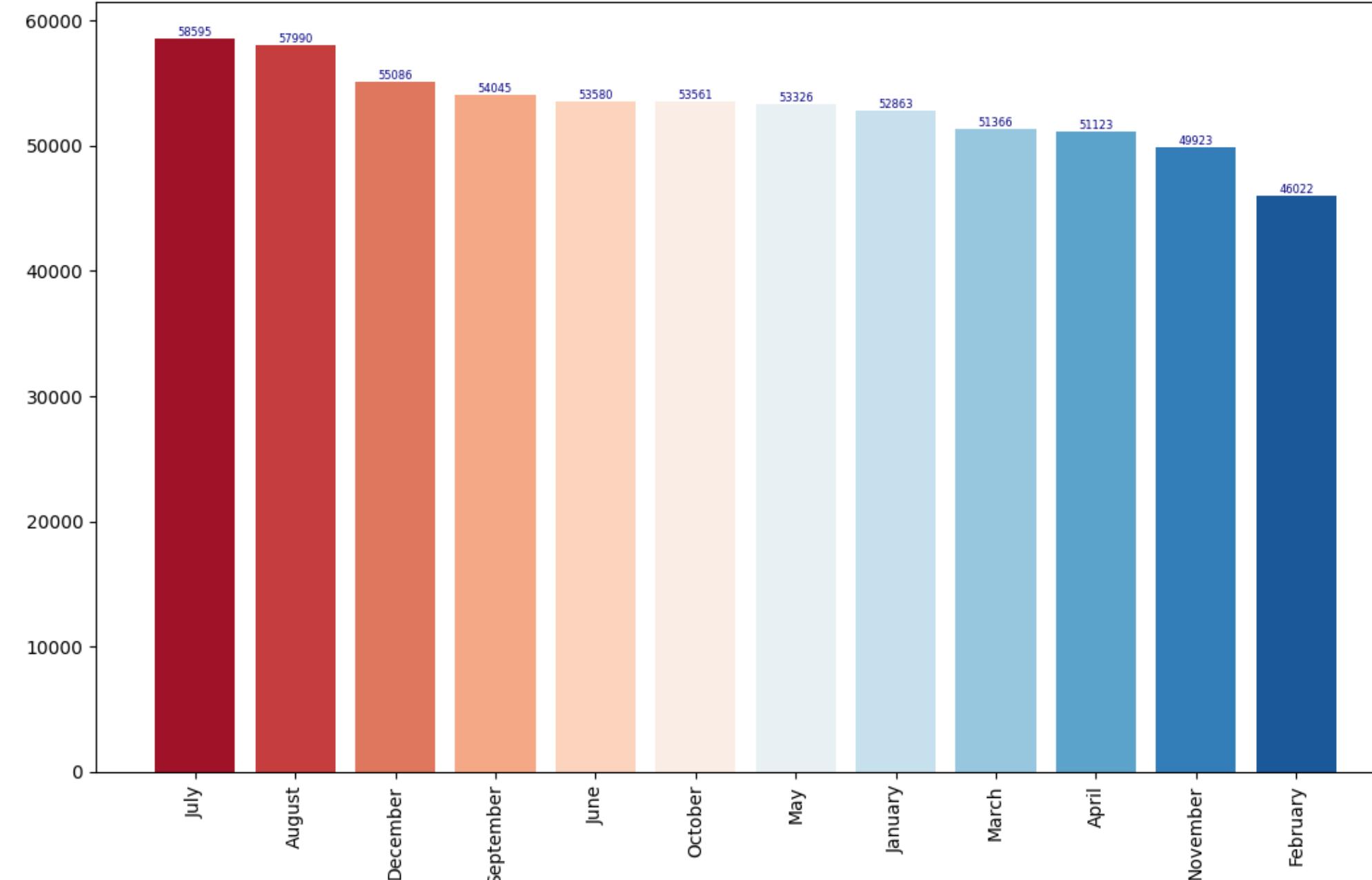


PERPETRATORS AND VICTIMS' AGES



HOMICIDES BY MONTHS

Count Of Homicide Cases Over The Months in US



August - 5756



August - 9242

July - 5717

July - 9171

September - 5348

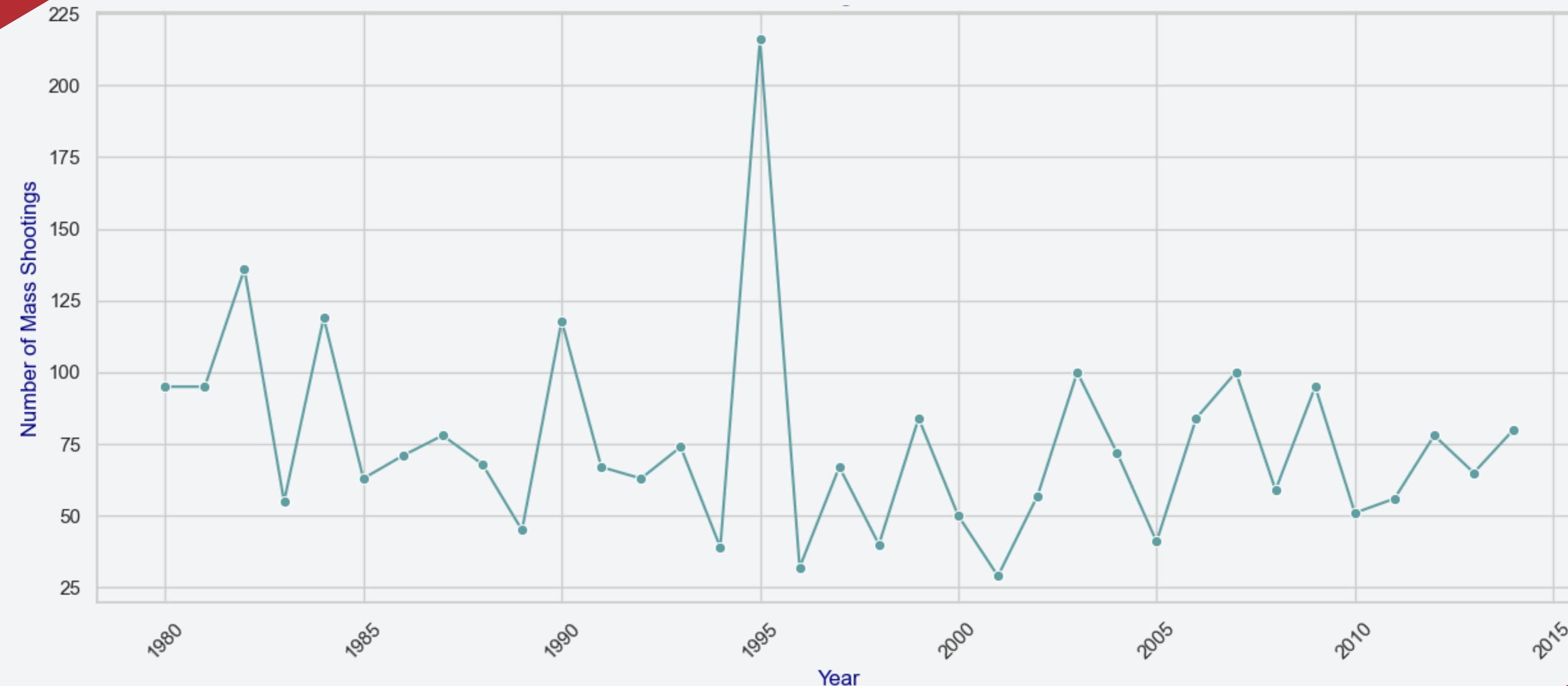
October - 8589

1

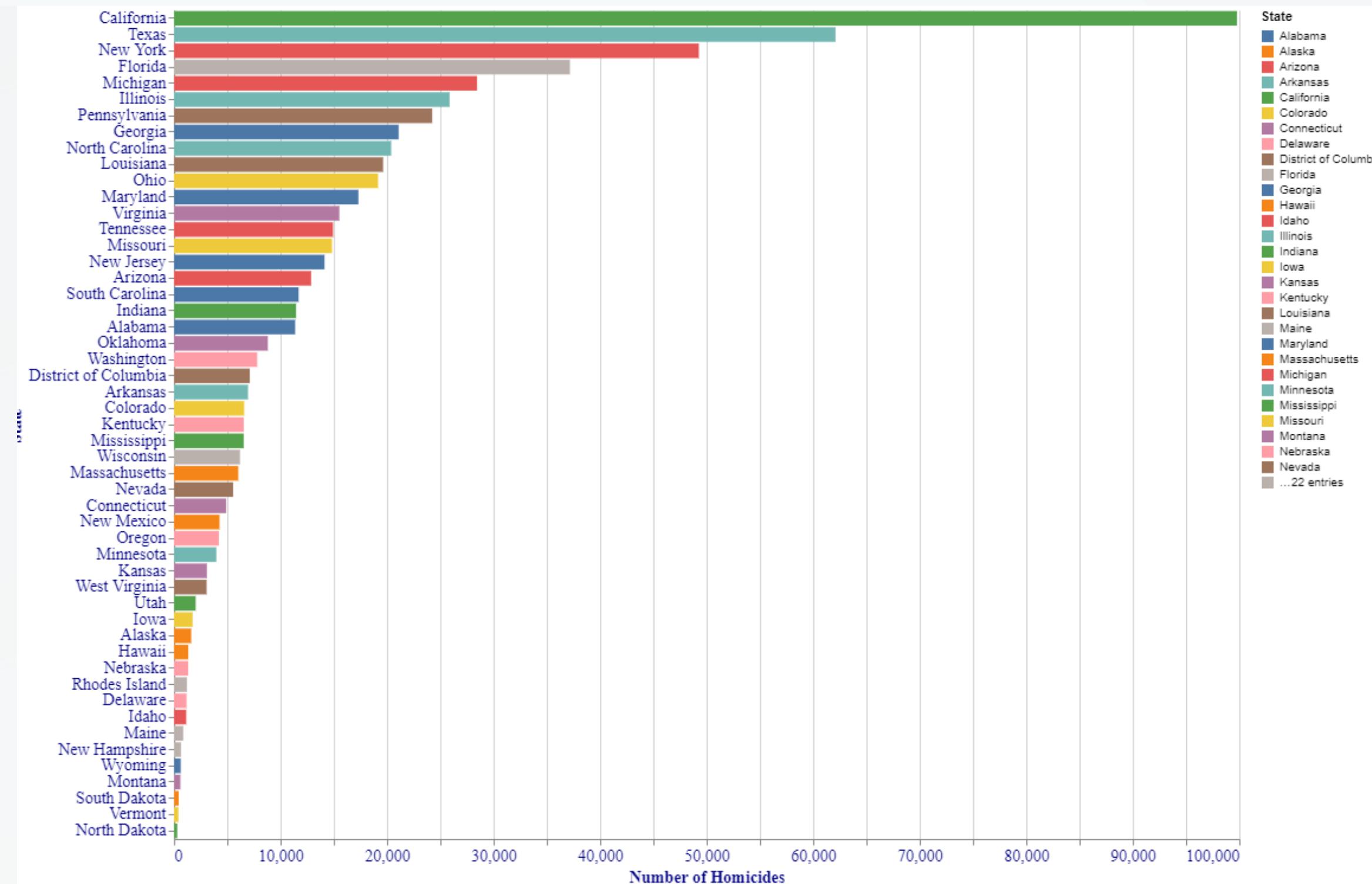
2

3

MASS SHOOTING OVER TIME

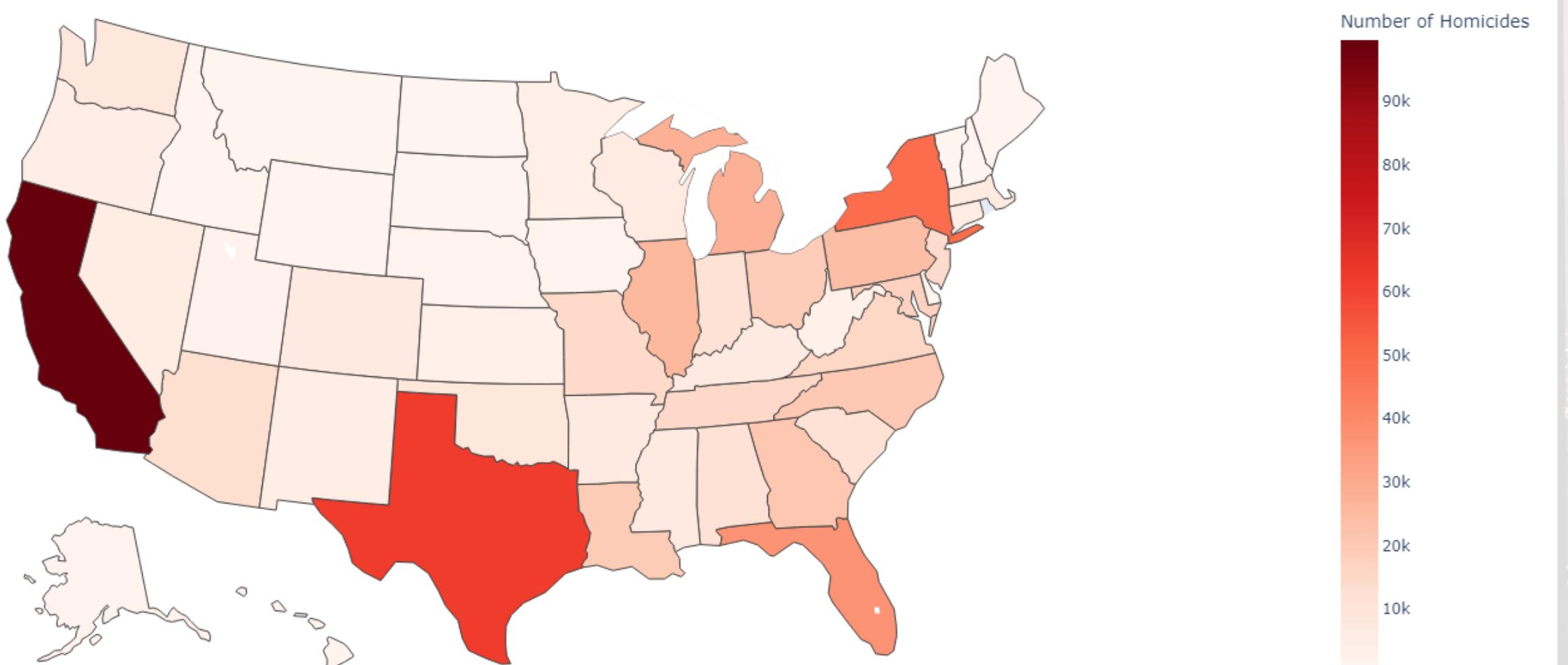


STATE DISTRIBUTION OF HOMICIDES

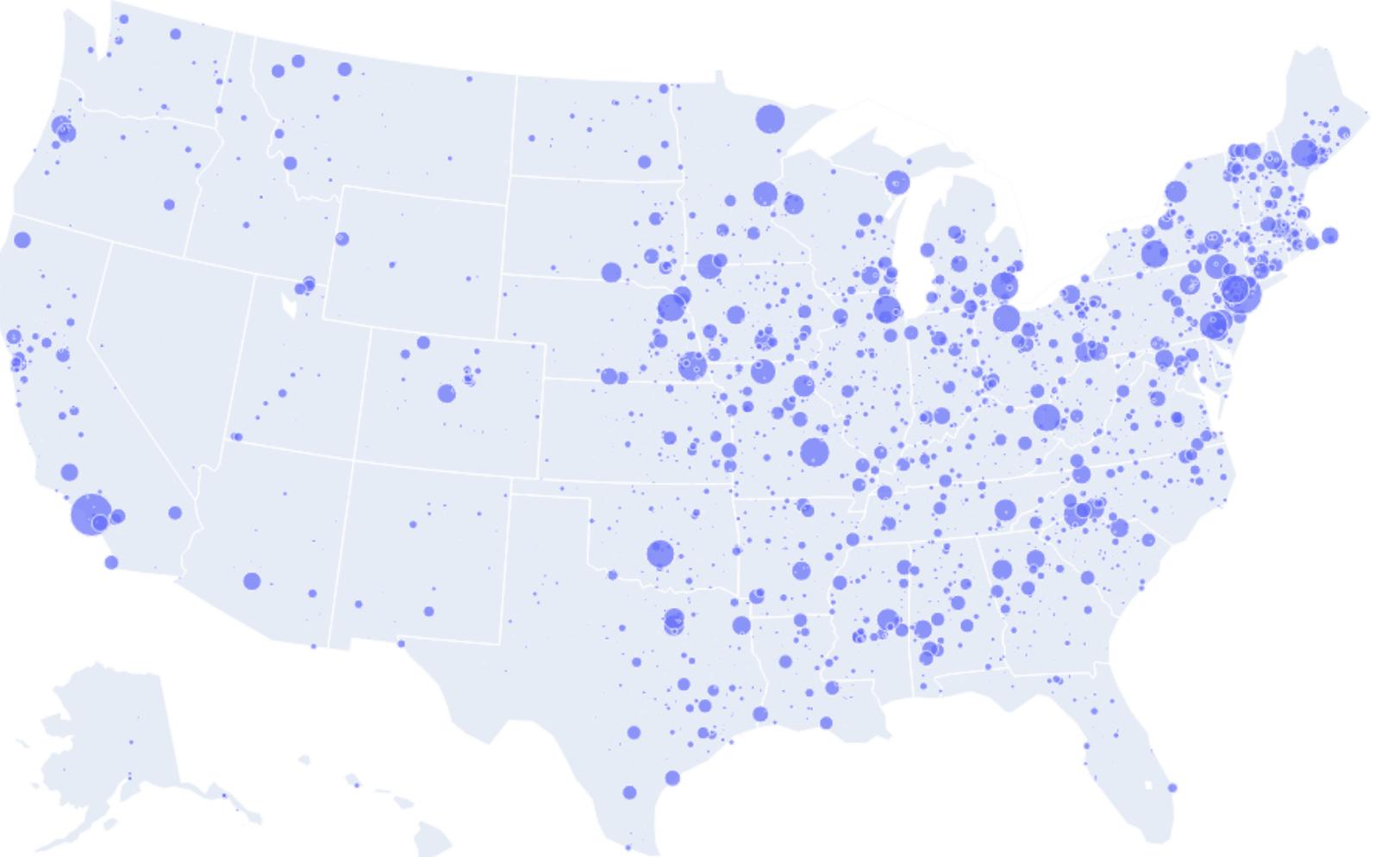


THAT CAN BE ALSO
INTERPRETED AS...

HEATMAP OF HOMICIDES



CITY-WISE DISTRIBUTION OF HOMICIDES



Weapons

Handgun

Rifle

Knife

Fire

Shotgun

Firearm

Blunt Object

Unknown

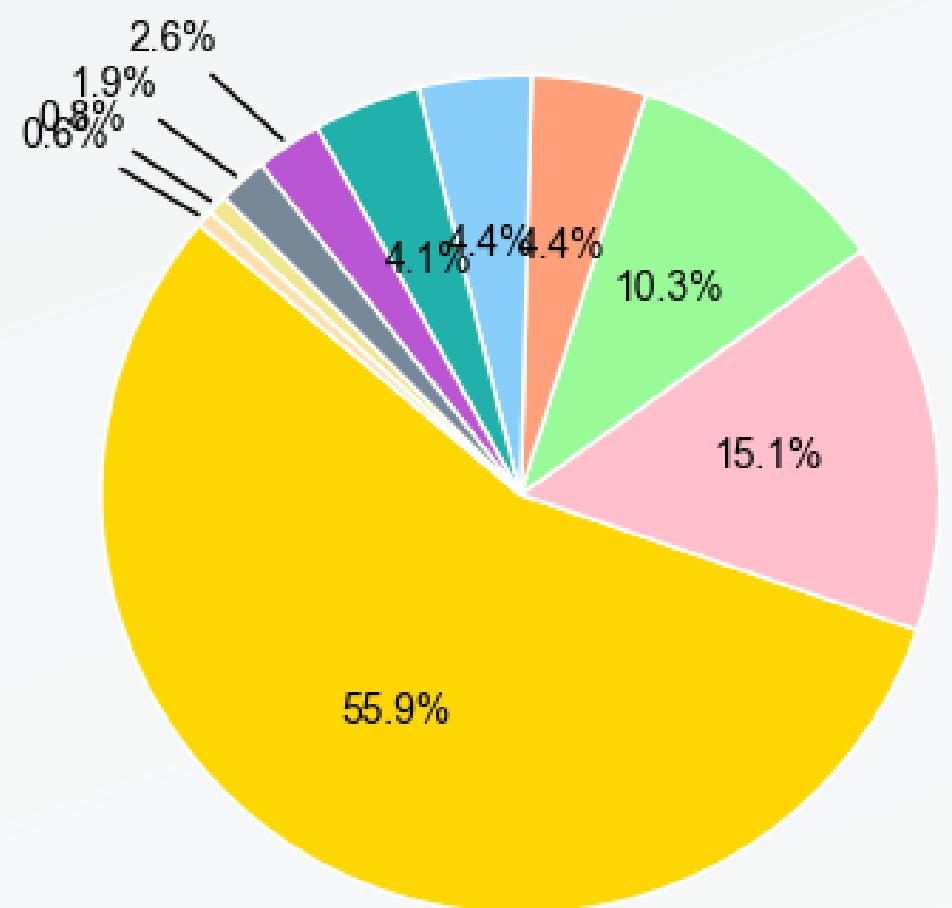
Suffocation

Strangulation

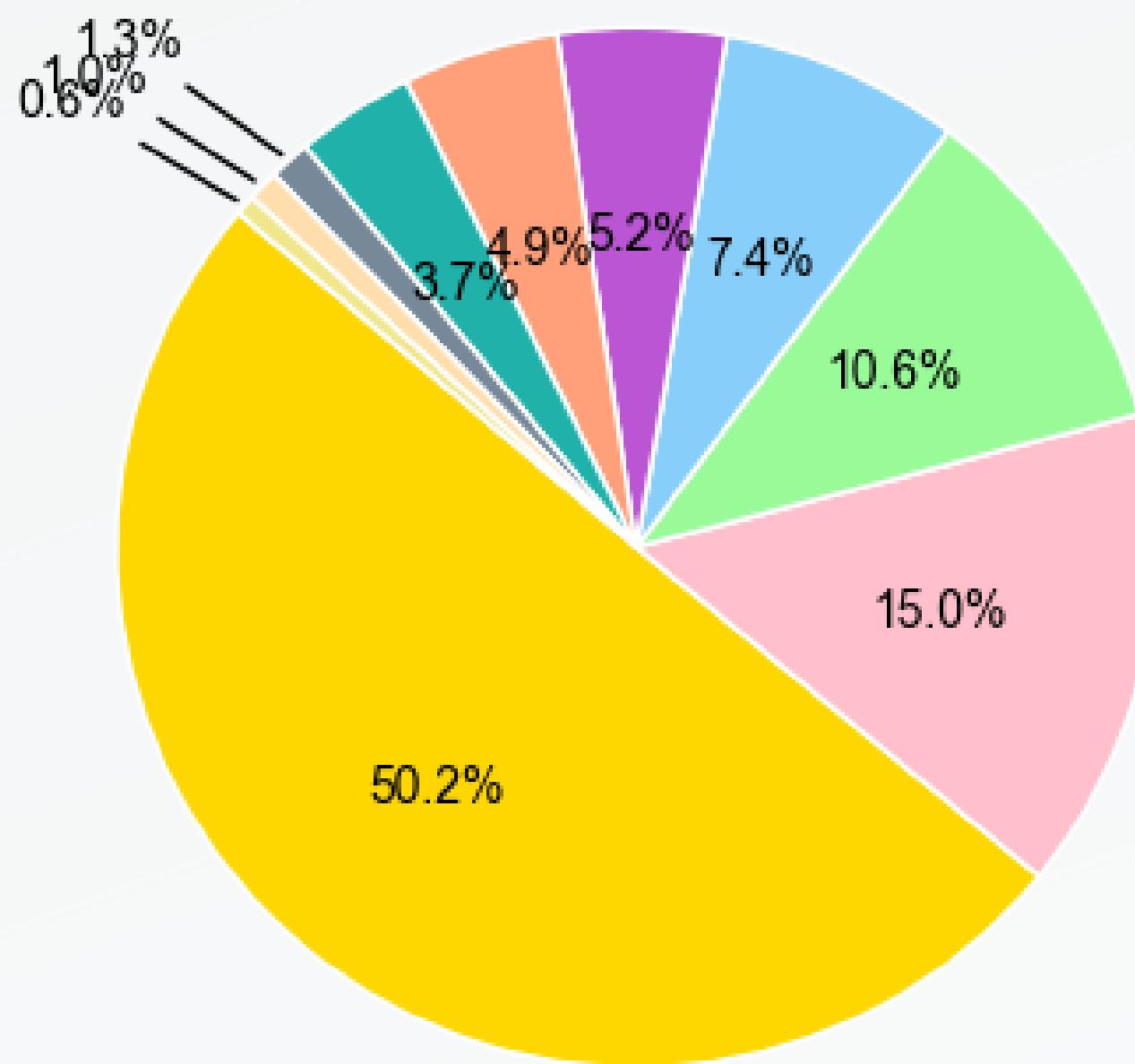
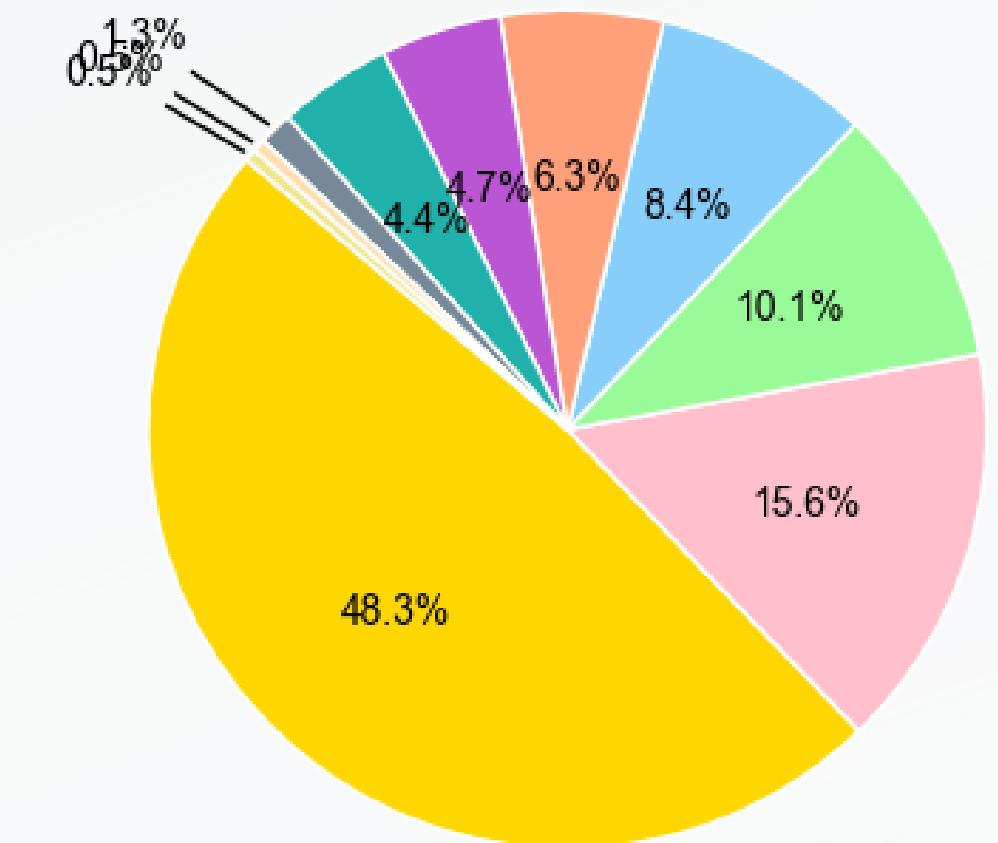
MOST USED WEAPONS

USA

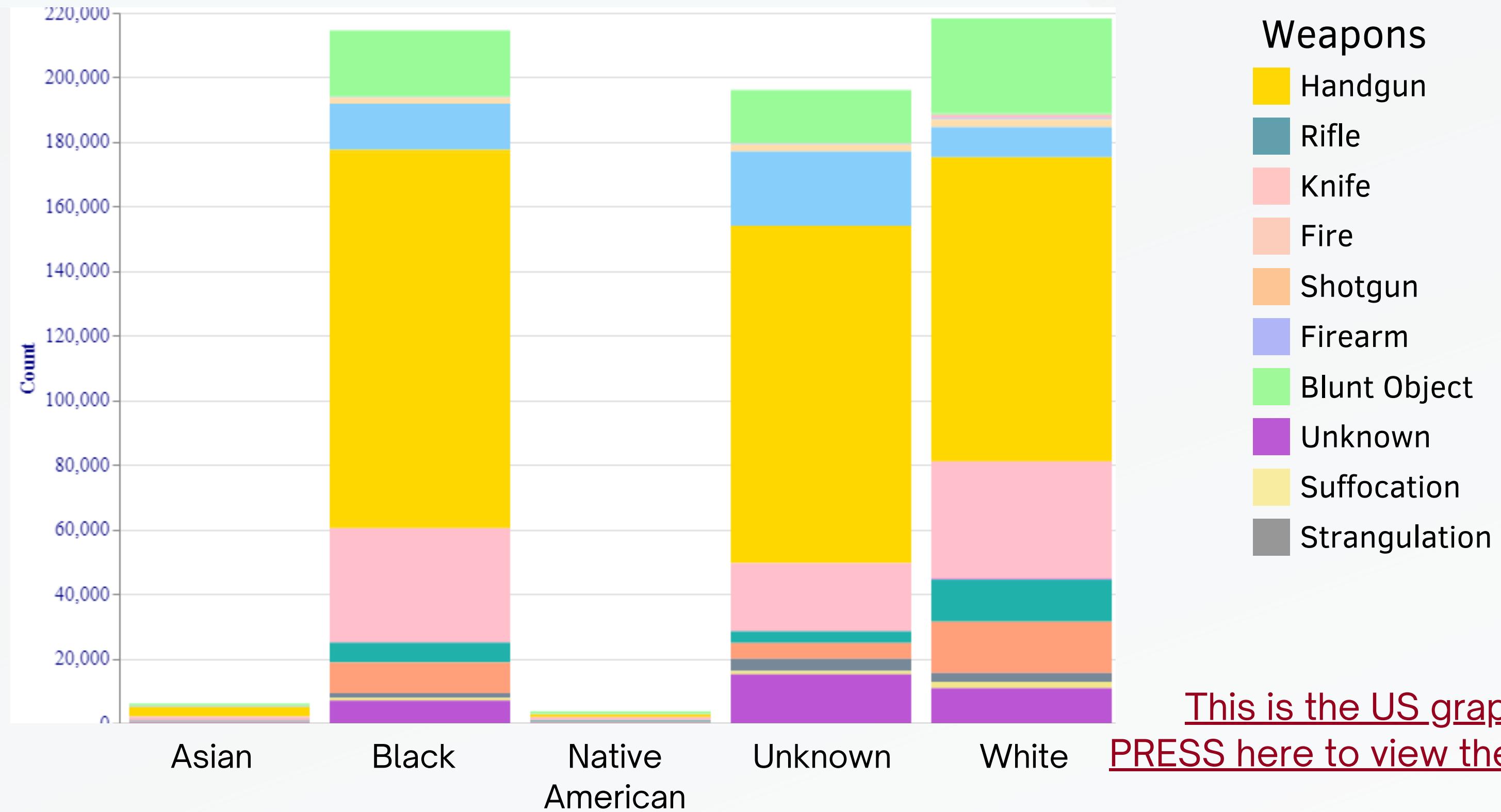
CALIFORNIA



TEXAS



PREFERRED WEAPON BY RACE



This is the US graph.

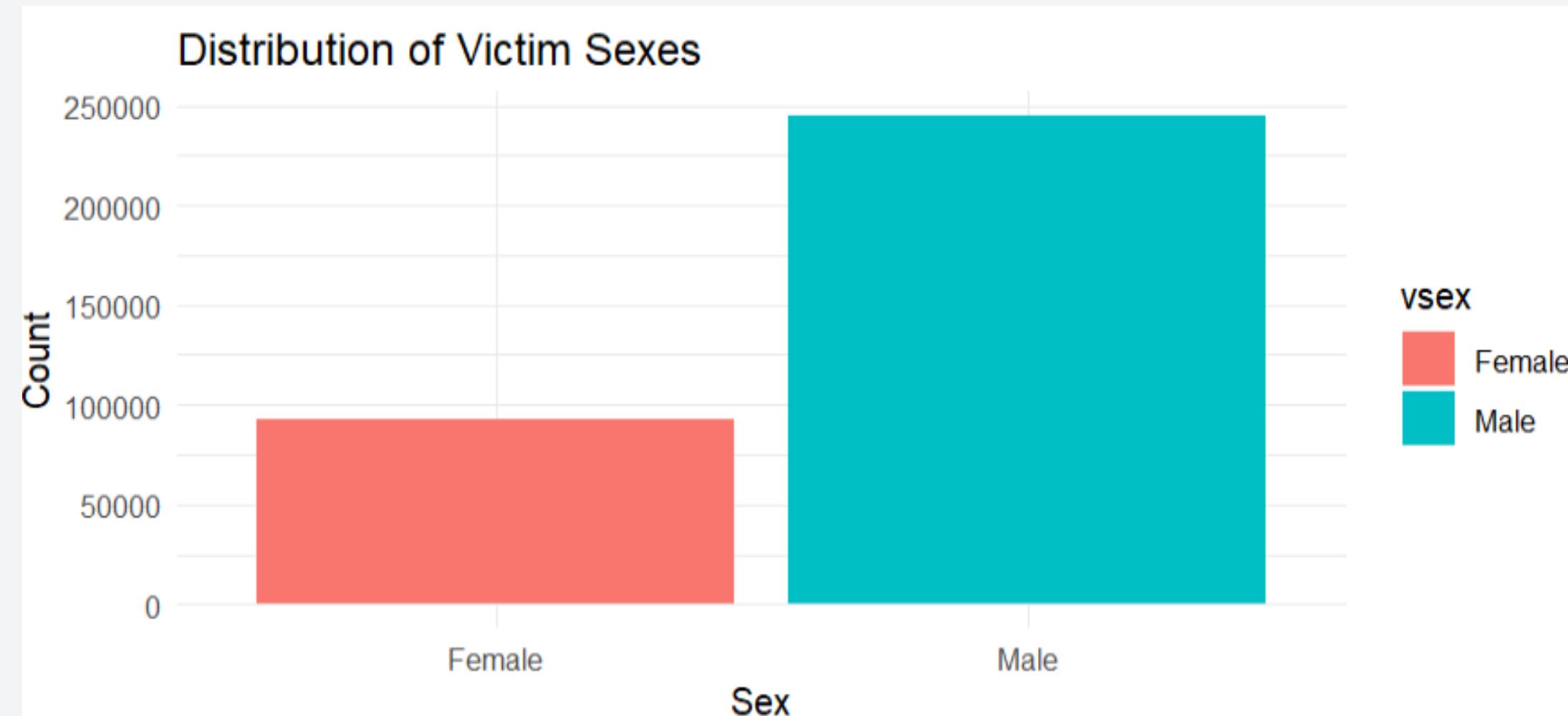
[PRESS here to view the others](#)



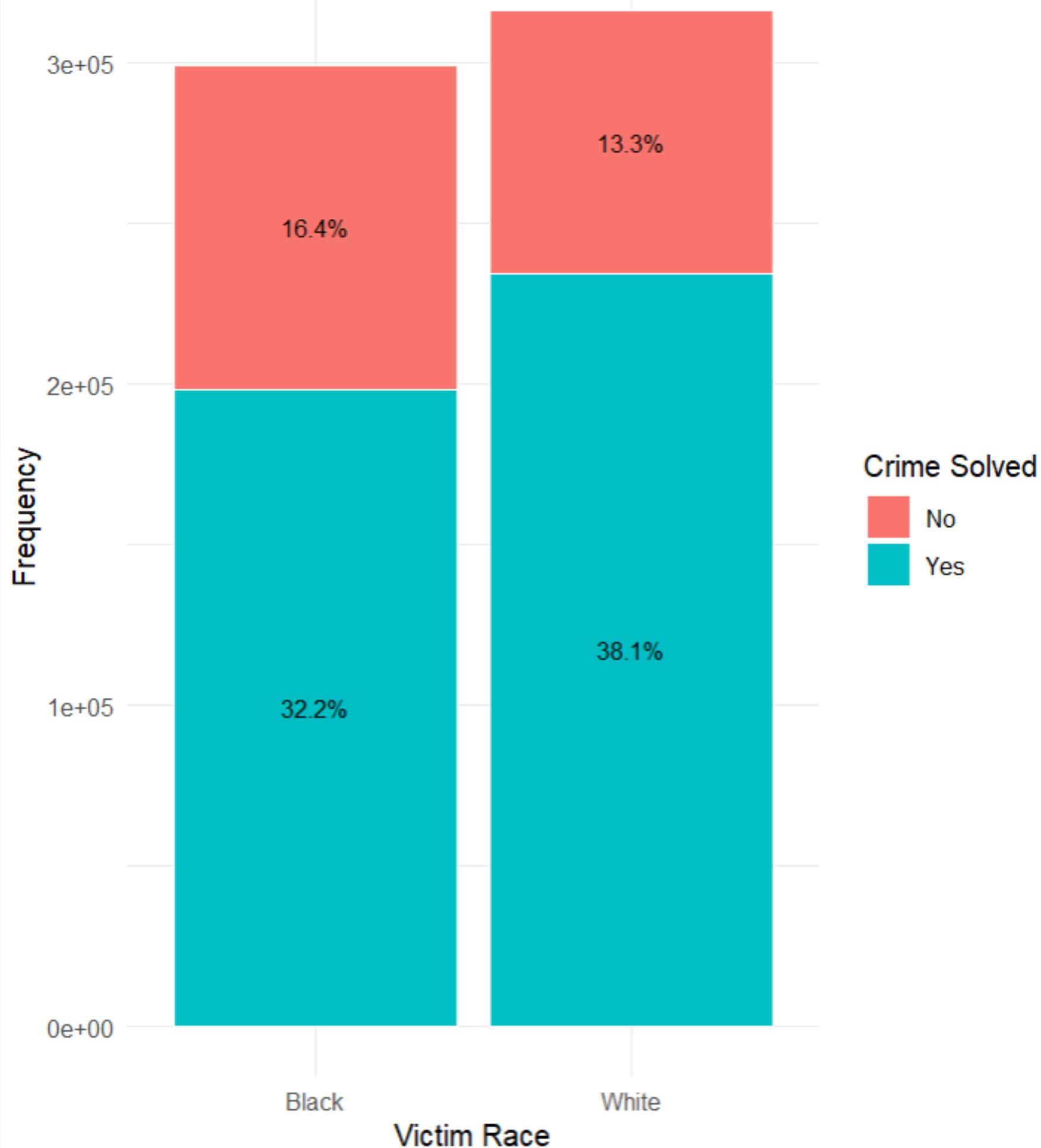
VICTIM SEX

MALE - FEMALE

the data suggests that
most of homicides
victims are Male



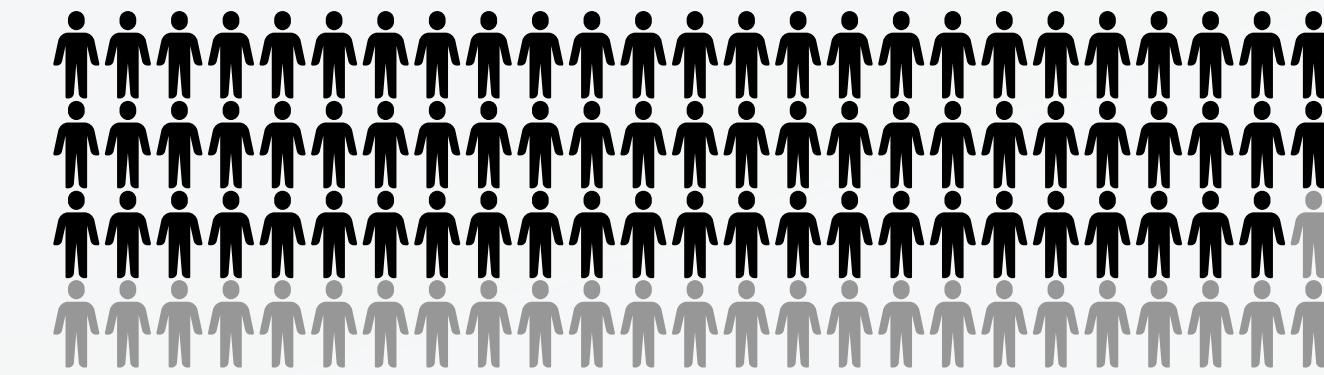
Bar Plot of Crime Solved by Victim Race



CRIME DETECTION RATE

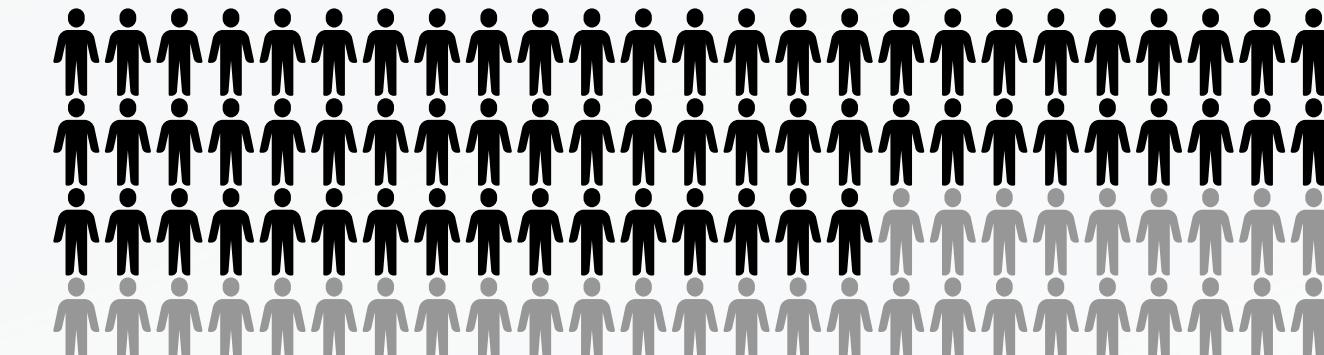
IN WHITE PEOPLE

74.1%



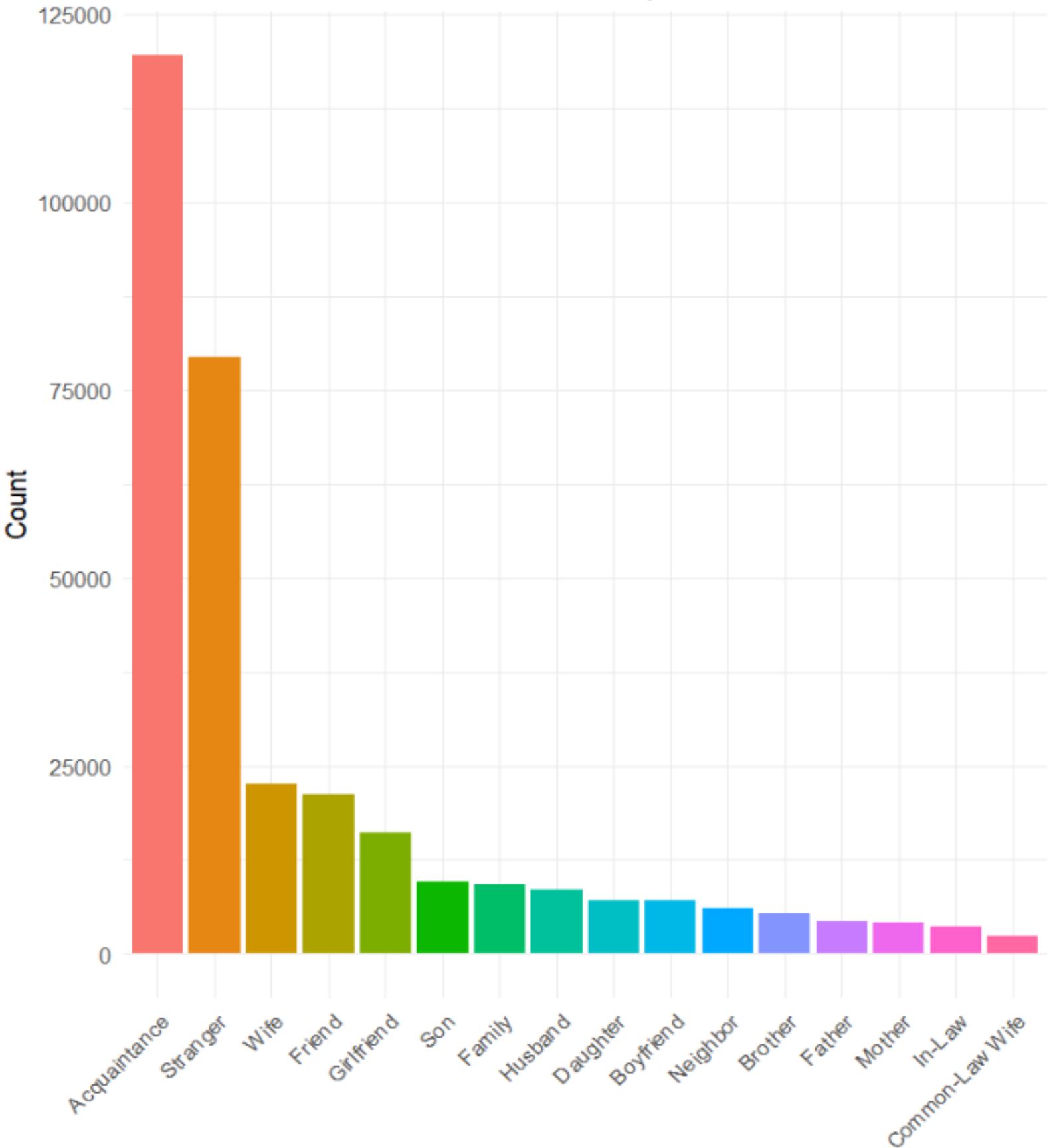
IN BLACK PEOPLE

66.2%

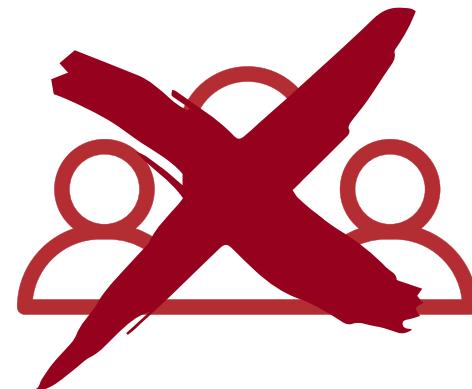


WHICH IS THE MOST DANGEROUS RELATIONSHIP?

Distribution of Homicide Relationship



- relationship
- Acquaintance
- Stranger
- Wife
- Friend
- Girlfriend
- Son
- Family
- Husband
- Daughter
- Boyfriend
- Neighbor
- Brother
- Father
- Mother
- In-Law
- Common-Law Wife



Acquaintance

Most of times people are motivated to kill



Stranger

Robbery, theft, rape could be the reason



Wife

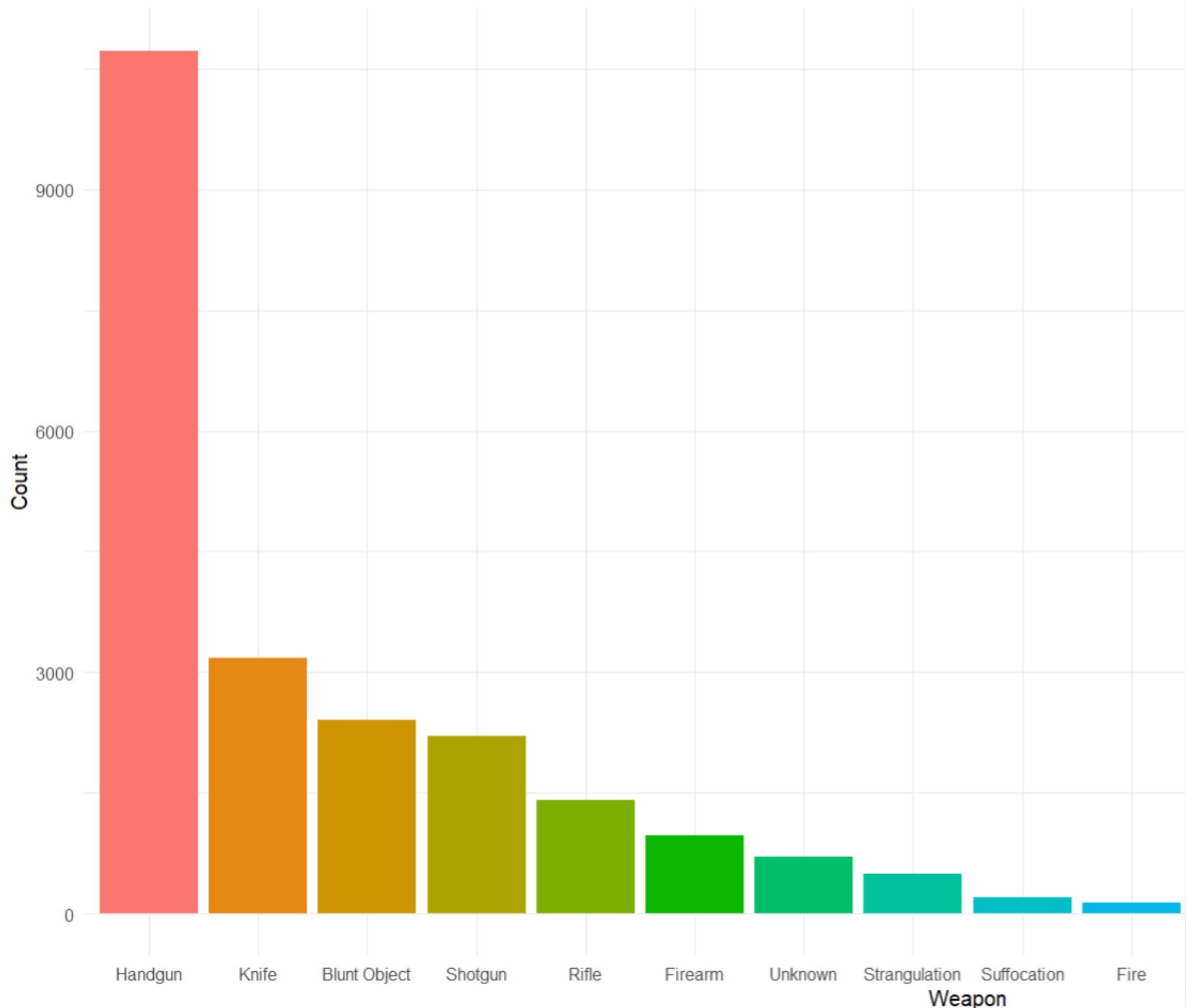
Domestic violence, emotional disputes



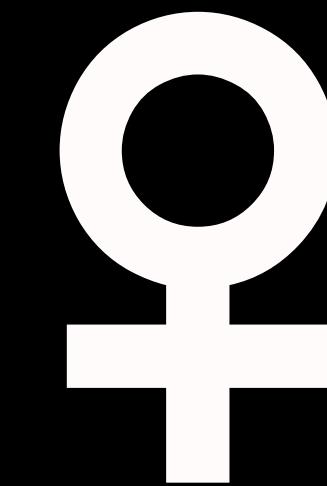
Friends

Your friends piss you off sometimes.

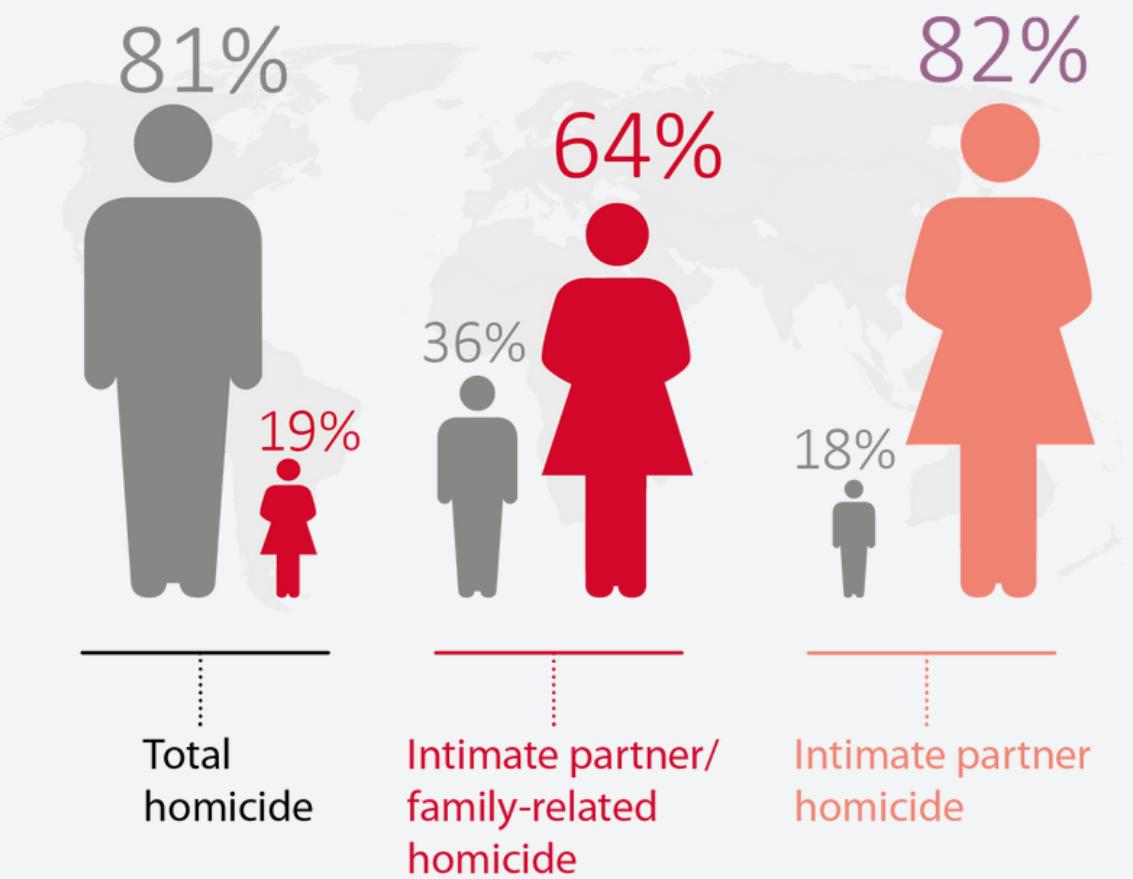
Which Weapon Kills Most Wives



WIFE VICTIMS



The most used weapon indoors is the handgun



Prediction Model

1

Select only relevant state

```
dataset = homicide[homicide$state %in% c('California', 'Texas', 'New York'),]  
summary(dataset$state)
```

2

Select only relevant variable

```
df = dataset %>% select(9:15)
```

3

Split into train and test set

```
index = sample(nrow(df), size = 0.8*nrow(df))  
train = df[index,]  
test = df[-index,]
```

4

Use glm() function

```
logall = glm(vsex ~ ., family = 'binomial', data = train)  
logall.pred = predict(logall, test, type = 'response')
```

Prediction Model (cont.)

5

Find optimal threshold and ROC curve

```
#find the optimal threshold for ROC curve
roc_curve = roc(test$vsex, logall.pred)
optimal_threshold = coords(roc_curve, "best", ret = "threshold")$threshold
test = test %>%
  mutate(pred = logall.pred)
test = test %>%
  mutate(pred = case_when(pred > optimal_threshold ~ 'Male', TRUE ~ 'Female'))
```

6

Check the accuracy

```
logall.accuracy = mean(test$vsex == test$pred)
cat("The accuracy is", round(logall.accuracy * 100, 2), "%\n")
> cat("The accuracy is", round(logall.accuracy * 100, 2), "%\n")
The accuracy is 87.76 %
```

THANKS FOR YOUR ATTENTION!

Edoardo Bollati

Claudia Cortese

Elena Tomasella

Zhao Wei Zhu

