

Understanding Homicide in the United States

*Carin Gan
Darius Lim
Matias Porras Paniagua*

Contents

1	Introduction	2
2	Summary of the Data	2
3	Data Cleaning and Preparation	3
3.1	Reduction of dataset to working variables:	3
3.2	Creating New Variables:	3
4	National Trends	5
5	State-wide Trends	8
6	Individual-level Trends	9
6.1	Categorizing Victims and Perpetrators based on Characteristics	9
6.2	Deep Dive of Perpetrators and Victims	9
6.3	Relationship between the Victim & Perpetrator	12
6.4	Weapon Preference Across Murderers	12
7	Conclusion	14

1 Introduction

The United States has long had the problem of high homicide rates. Despite being one of the most developed countries in the world, it has a homicide rate that rather resembles those of third-world nations. In 2017, the United Nations Office on Drugs and Crime reported that the US had a homicide rate of 5.3 per 100,000 inhabitants, which is higher than that of Cuba and North Korea.

In this study, we took a deep dive into the data to paint a better picture of the homicide situation in the United States. We first looked at overarching trends across the nation, then state-level patterns, and lastly statistics that concerned individual homicide perpetrators and victims.

The intent of this paper is not to make any new miraculous discovery on how to resolve the age-old problem of crime. Rather, it seeks to educate the reader on this prevalent societal problem and hopefully allow him/her to make informed discussions on the issue.

2 Summary of the Data

The Murder Accountability Project is the most complete database of homicides in the United States. The dataset includes murders from the FBI's Supplementary Homicide Report from 1976 to the present. Our dataset was obtained from Kaggle. This dataset includes the age, race, sex, ethnicity of victims and perpetrators, in addition to the relationship between the victim and perpetrator and weapon used from the FBI and FOIA requests. The dataset consists of more than 600,000 reported crimes with 24 variables. The dataset contained the following variables:

- Record ID: Unique identifier associated with every case
- Agency Code: Code identifying the agency that reported the crime
- Agency Name: Name of the agency that reported the crime
- Agency Type: Type of agency that reported the crime
- City: City where the crime was committed
- State: State where the crime was committed
- Year: Year in which the crime was committed
- Month: Month in which the crime was committed
- Incident: Number of cases of a given agency in a given month and year
- Crime Type: Type of crime can be either "Murder or manslaughter" or "murder by negligence"
- Crime Solved: Whether the crime was solved or not
- Victim Sex: The sex of the victim
- Victim Age: The age of the victim
- Victim Race: The race of the victim
- Victim Ethnicity: The ethnicity of the victim
- Perpetrator Sex: The sex of the perpetrator
- Perpetrator Age: The age of the perpetrator
- Perpetrator Race: The race of the perpetrator
- Perpetrator Ethnicity: The ethnicity of the perpetrator
- Relationship: The relationship between the victim and the perpetrator
- Weapon: The weapon used by the perpetrator
- Victim Count: Number of victim count seen by the agency
- Perpetrator Count: Number of perpetrators seen by the agency
- Record Source: Whether the report came from the FBI or FIOA

3 Data Cleaning and Preparation

3.1 Reduction of dataset to working variables:

- 1) Variables with no degrees of freedom were removed since there was a unique value for each observation:
 - Record ID
- 2) We also removed variables that represent less than 5% of the observations. This includes:
 - Victim Sex = “Unknown”
 - Crime Type = “Manslaughter by Negligence”
- 3) We also removed observations with missing values.
- 4) We also removed variables with no relevance to a particular crime:
 - Victim Count
 - Perpetrator Count
 - Incident
- 5) Removal of redundant variables (very high correlation with the city):
 - Agency Code
 - Agency Name
- 6) We decided to remove variables with one level since they have no variance:
 - Crime Type: One level (“Murder or Manslaughter”)

3.2 Creating New Variables:

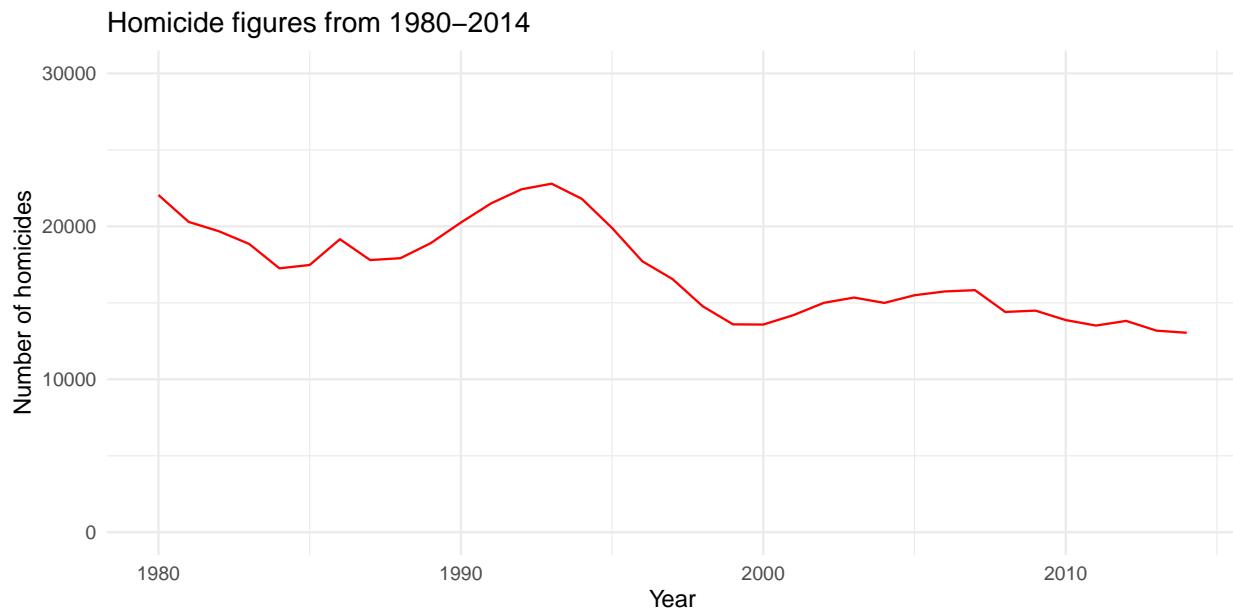
We decided to create new features to explore more in-depth the relationships between the murderer and the victim and the methods in which the homicide was conducted:

- 1) Firearm: Whether the perpetrator used a firearm or not. This is a binary variable.
 - Firearm: 1 if a firearm, shotgun, gun, handgun or rifle; 0 otherwise
- 2) Broader_Relationship: Categorizing the relationship between the perpetrator and victim at a broad level. This variable has 8 levels.
 - Core Family: Common-Law Husband, Daughter, Father, Son, Brother, Common-Law Wife, Mother, Sister, Wife
 - Dating: Boyfriend, Girlfriend, Boyfriend/Girlfriend
 - Extended Family: In-Law, Stepfather, Stepson, Stepdaughter, Stepmother
 - Ex-Partner: Ex-Wife, Ex-Husband
 - Work Related: Employer, Employee
 - Acquaintance: Neighbor, Friend, Acquaintance
 - Stranger: Stranger

- Unknown
- 3) Broad_Relationship: Categorizing the relationship between the perpetrator and victim at a less granular level. This variable has 3 levels.
- Family: Common-Law Husband, Daughter, Father, Brother, Son, Common-Law Wife, Husband, Mother, Sister, Wife, In-Law, Stepfather, Stepson, Stepdaughter, Stepmother, Ex-Husband, Ex-Wife, Family
 - Non-Family: Boyfriend/Girlfriend, Boyfriend, Girlfriend, Employer, Employee, Acquaintance, Neighbor, Friend, Stranger
 - Unknown
- 4) Region: Grouping 51 States into the following regions.
- Northeast: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, District of Columbia, Pennsylvania
 - Midwest: Ohio, Michigan, Indiana, Wisconsin, Illinois, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas
 - South: Delaware, Maryland, Virginia, West Virginia, Kentucky, North Carolina, South Carolina, Tennessee, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana, Texas, Oklahoma
 - West: Alaska, California, Oregon, Hawaii, Washington, Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah

4 National Trends

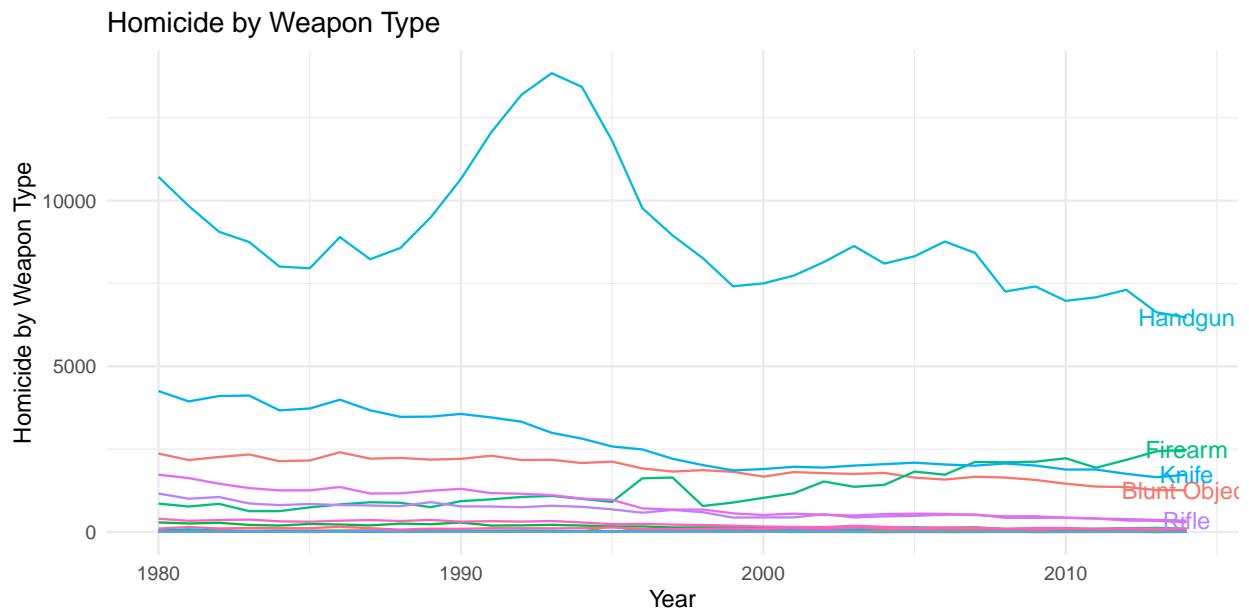
The first thing to analyze is how homicide figures have changed over time. Looking at the graph below, homicide numbers have generally decreased from 1980-2014.



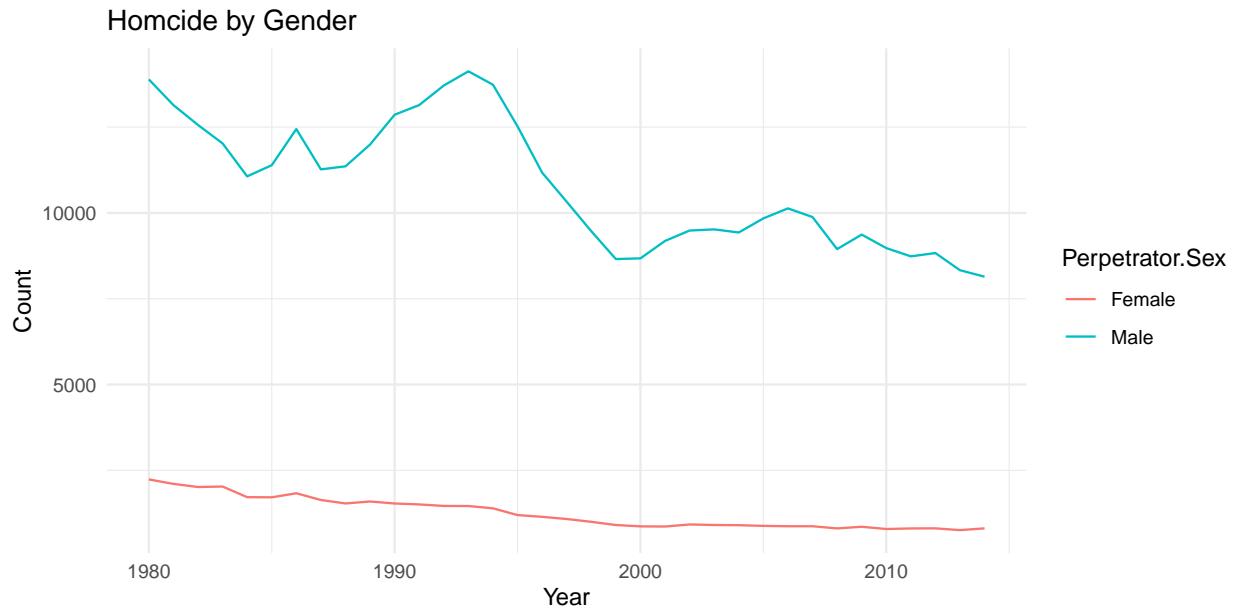
What is definitely noteworthy is that while there is a downward trajectory on the whole, there was a sudden spike in numbers in the early 1990s before they started to fall once again in the latter half of the decade.

This begs the question, what caused the spike in the 1990s?

As we see from the graph below, at the same time there was a spike in homicide crimes, there was also a large increase in the number of Handgun homicides. Other methods of homicide rates, in the meantime, remained relatively constant. Without needing to perform a statistical analysis, it is clear to see the two measures are highly correlated.

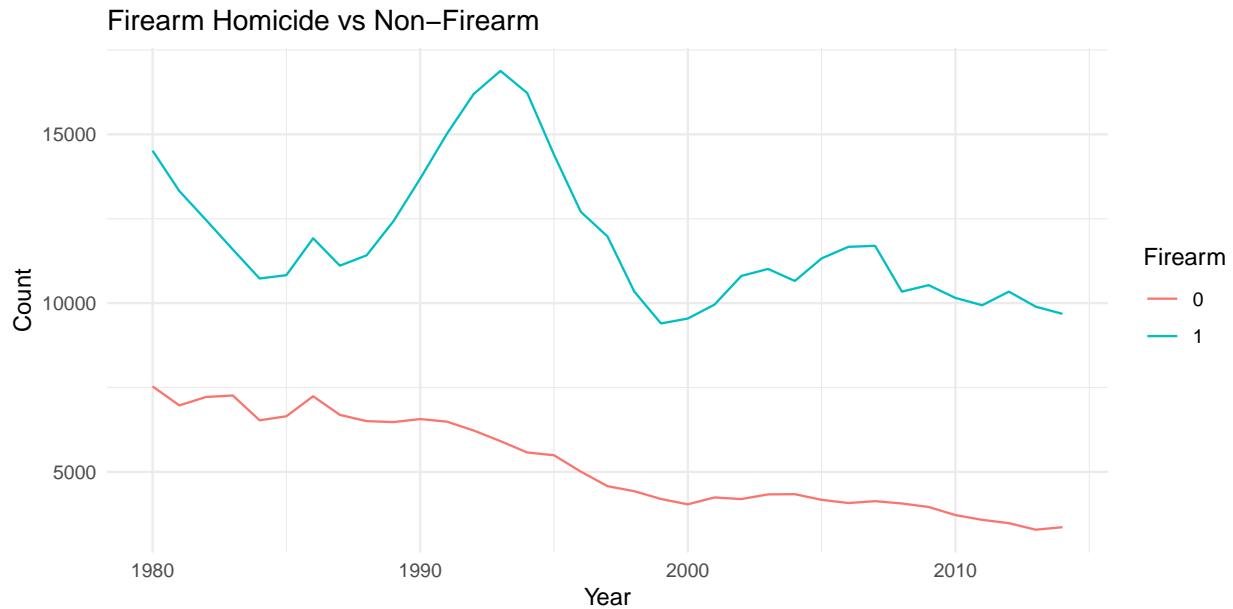


Looking at national homicide numbers by Gender, we also see the similar trend, in that Male homicide perpetrators spiked in the 1990s, while Female perpetrators have decreased without pause.



These trends, in fact, have been a hot topic of debate for academics. The most popular theory combines all three trends and comes to the conclusion that the spike was caused by the sudden appearance of crack cocaine in big cities around 1985. This lucrative opportunity was seized on young males across the country and dramatically increased the need for and use of guns by male drug dealers to defend their businesses. (See Blumstein, Rivara & Rosenfeld (2000))

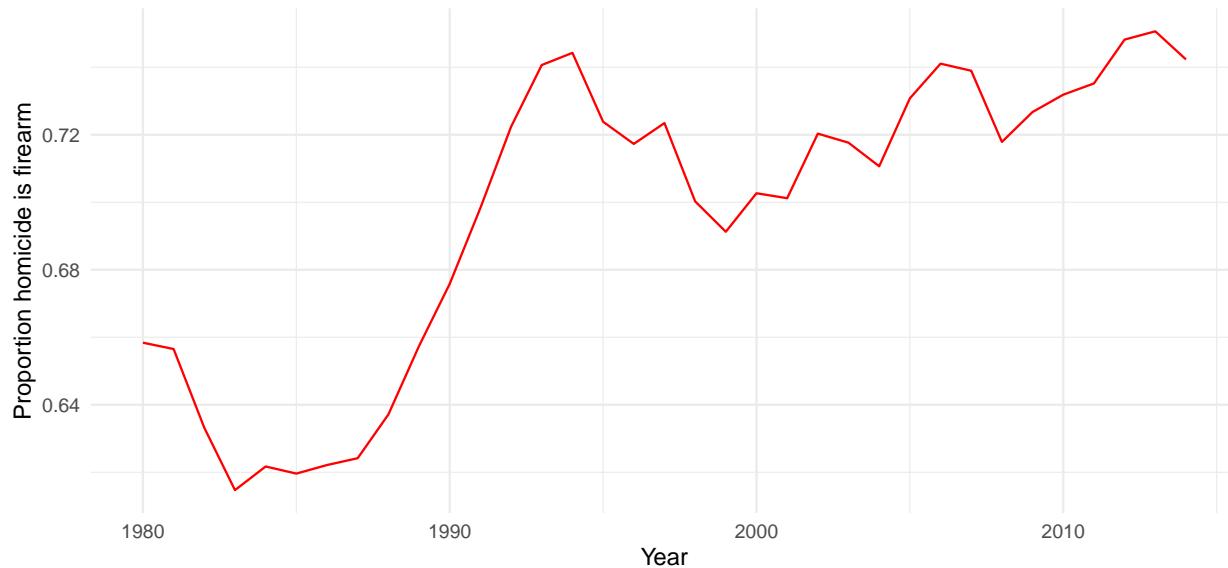
Since gun control remains a hot topic in the present day, we decided to look further into it. If we group all the means of weapon used broadly into “Firearm” vs “Non-Firearm” we find, unsurprisingly, the same spike in Homicide by Firearms in the 1990s. For Non-Firearms, the numbers continuously decreased across the entire timeframe of 34 years.



And while it is a great sight to see that homicide numbers have decreased drastically over the years, the

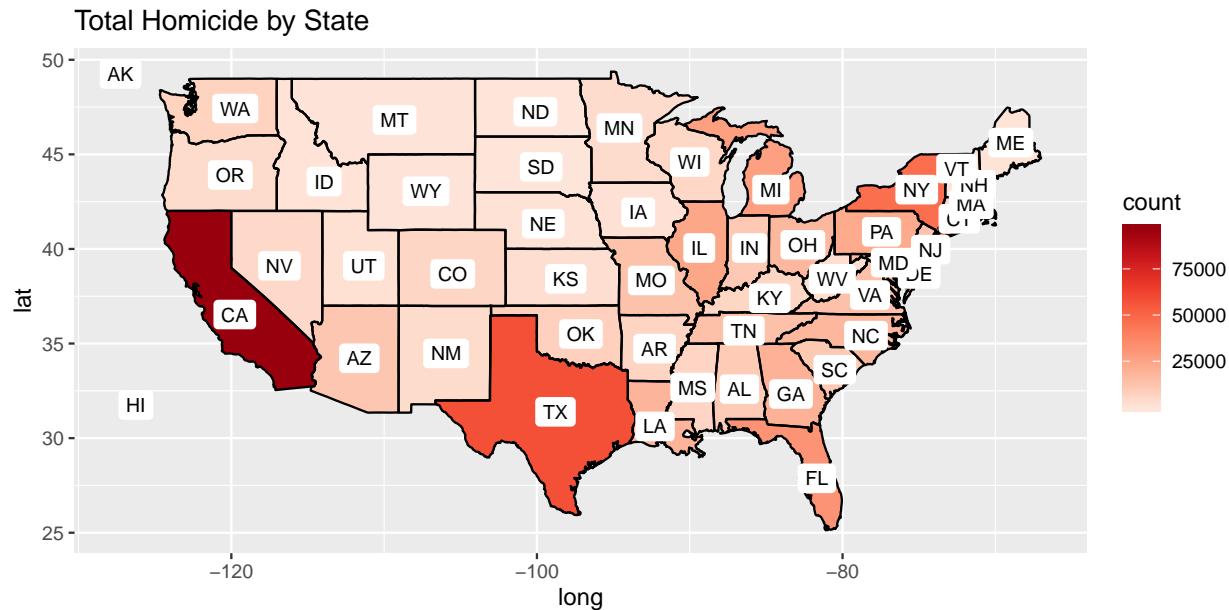
graph below suggests that if you were to fall victim to homicide in the present day, the chances of it being caused by a firearm are incredibly high.

Proportion of Homicides using Firearms



5 State-wide Trends

Having examined national statistics, we now look into state-level statistics. On first glance, California has a shockingly high rate of homicides compared to the rest of the country.

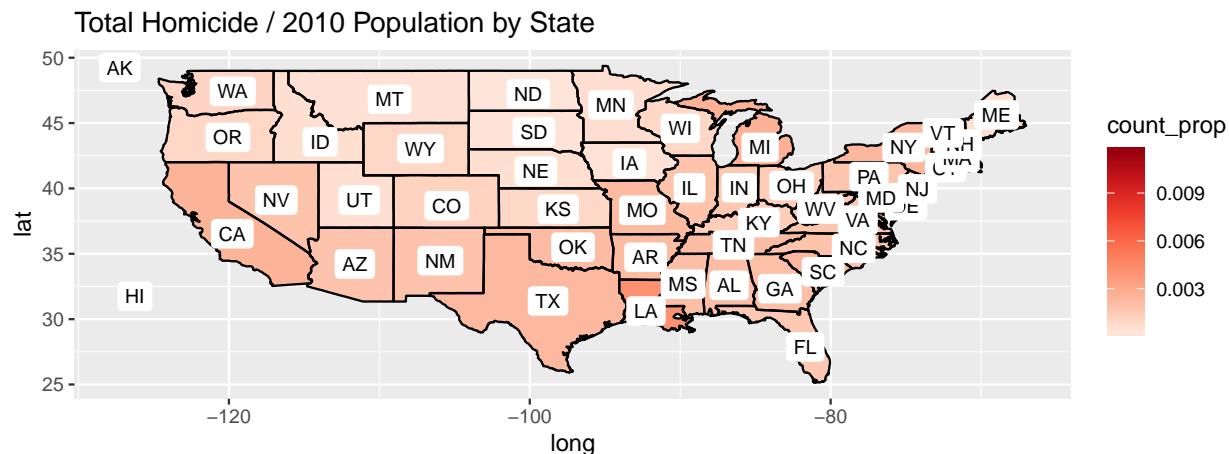


However, the statistic that would probably be more relevant is homicide number / population for every state.

To account for population number, we use data from the US Census Bureau. The 2010 population numbers will be used as a proxy for population number.

Although the exact percentage of total homicide / population per year will be inaccurate, what we are interested in is the relative percentages across state, and assuming population growth is constant across states this should be a good proxy.

As seen from the graph below, total homicide per population is actually relatively similar across the country. If anything, it seems that the ratio is the highest in the south, with Louisiana unfortunately having the highest rate of 0.4%.



6 Individual-level Trends

6.1 Categorizing Victims and Perpetrators based on Characteristics

Lastly, we also want to analyze how homicide trends have changed at the individual level. For individuals, we consider both the **Victim** and the **Perpetrator**.

The dataset allows us to analyze individuals based on 4 characteristics:

- Age
- Race
- Gender
- Ethnicity

Because Race, Gender and Ethnicity are categorical, we found it difficult to use traditional clustering methods. Instead, we broadly categorized individuals into groups based on their characteristics, with each individual having three of seven possible traits:

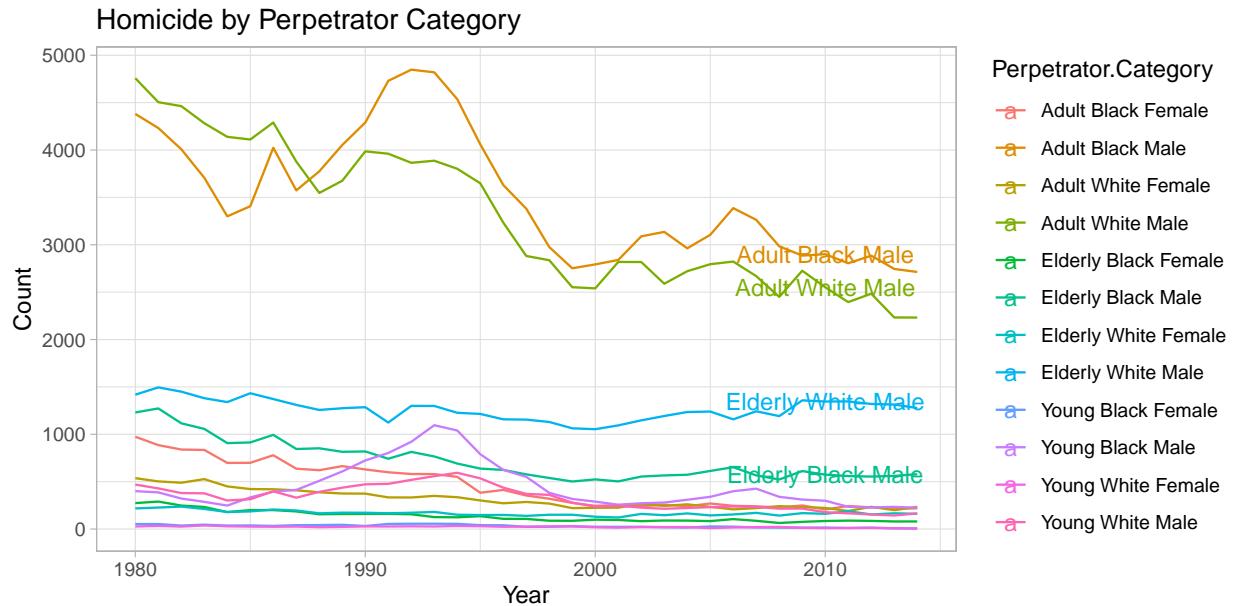
- Age: Young (0-18), Adult (19-40) & Elderly (>40)
- Race: Black & White (Other races were not considered as they consisted of too few datapoints)
- Gender: Male & Female

The result is 12 categories each for Victim and Perpetrator. Examples of categories are *Young Black Male* and *Elderly White Female*.

(Ethnicity was not considered due to there being too many “Unknown” values)

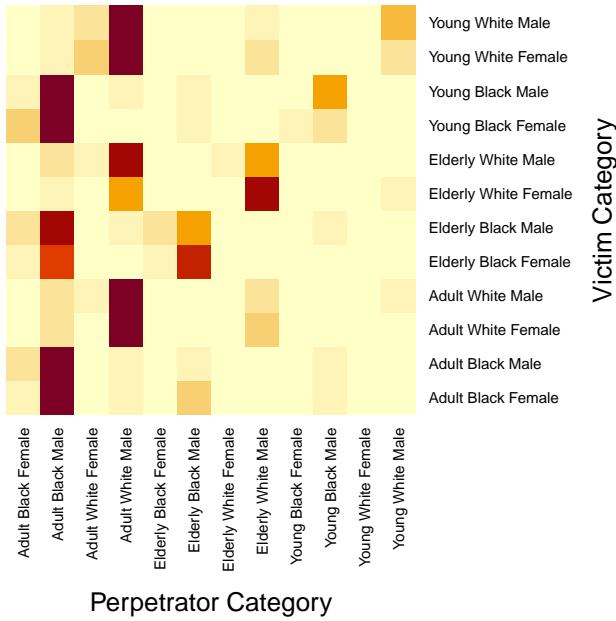
6.2 Deep Dive of Perpetrators and Victims

Plotting Homicide Numbers by Perpetrator Category, we find the following result:



In looking at the Homicide numbers in the 1990s, Adult White Males' numbers are relatively constant. Rather, it is Adult Black Males and Young Black Males whose numbers spiked significantly, which suggests they the African American population was most influenced by the rise of crack cocaine in the 90s.

Next, we analyze the total number of counts of homicide committed categories of Perpetrators against categories of Victims. Below is a heatmap of Perpetrators' categories by Victim Categories



If we look at the top categories of perpetrators and victims:

Table 1: Top Perpetrator-Victim Homicide Counts

Perpetrator	Victim	Count
Adult Black Male	Adult Black Male	63583
Adult White Male	Adult White Male	52043
Adult White Male	Elderly White Male	19970
Adult White Male	Adult Black Female	17120
Adult Black Male	Adult Black Female	13226

We find two interesting observations:

- 1) Crime doesn't seem to cross racial lines.

We hypothesize that this is because people are likely to be murdered by others in their neighborhood, and neighborhoods are usually dominated by a single race.

- 2) There might be a possible correlation between perpetrator age and victim age

This is based purely on the fact that Adults perpetrators are responsible for the killing of most Adults victims.

That is not a comprehensive conclusion in itself, hence we will conduct statistical analysis based on Perpetrator Age and Victim Age.

First, we fit a linear regression of Perpetrator Age on Victim Age. The summary is as follows below:

```
##
```

```

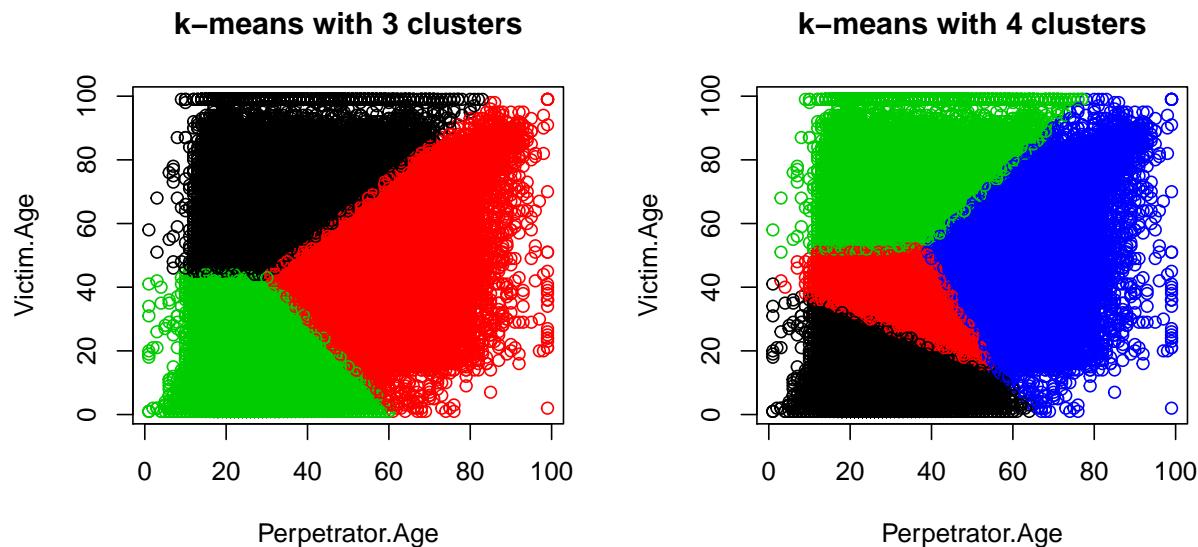
## Call:
## lm(formula = Perpetrator.Age ~ Victim.Age, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -173.373  -8.894  -2.894   6.270  73.013 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 2.568e+01 3.672e-02 699.4   <2e-16 ***
## Victim.Age  1.530e-01 8.986e-04 170.3   <2e-16 ***  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 12.38 on 374072 degrees of freedom
## Multiple R-squared:  0.07192,    Adjusted R-squared:  0.07192 
## F-statistic: 2.899e+04 on 1 and 374072 DF,  p-value: < 2.2e-16

```

On first glance, the p-value of the Victim.Age variable is statistically significant at the 0.01 level. On the other hand, the R^2 value is incredibly low, at 0.07192. Victim.Age doesn't explain much of the variance of Perpetrator.Age, hence we will look to further analysis to examine this relationship further.

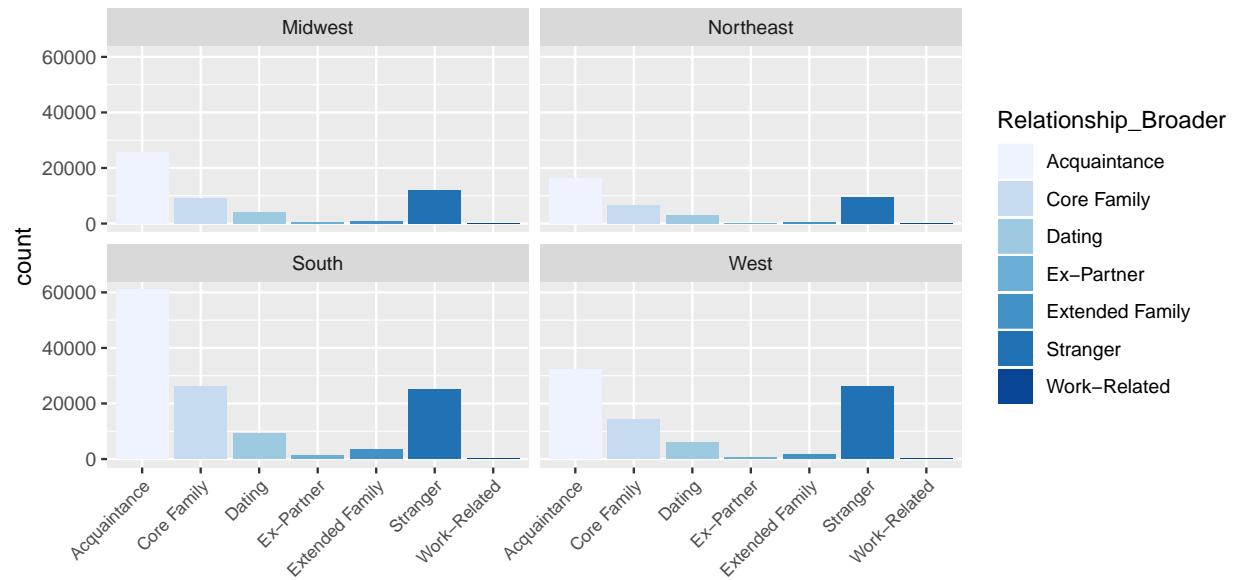
Next, we conduct a K-means unsupervised clustering analysis. K-means groups similar data points together and discovers underlying patterns; if there is an underlying pattern in Perpetrator.Age and Victim.Age, the K-means clustering should identify it.

Below, we have k-means with 3 clusters on the left, and 4 clusters on the right.



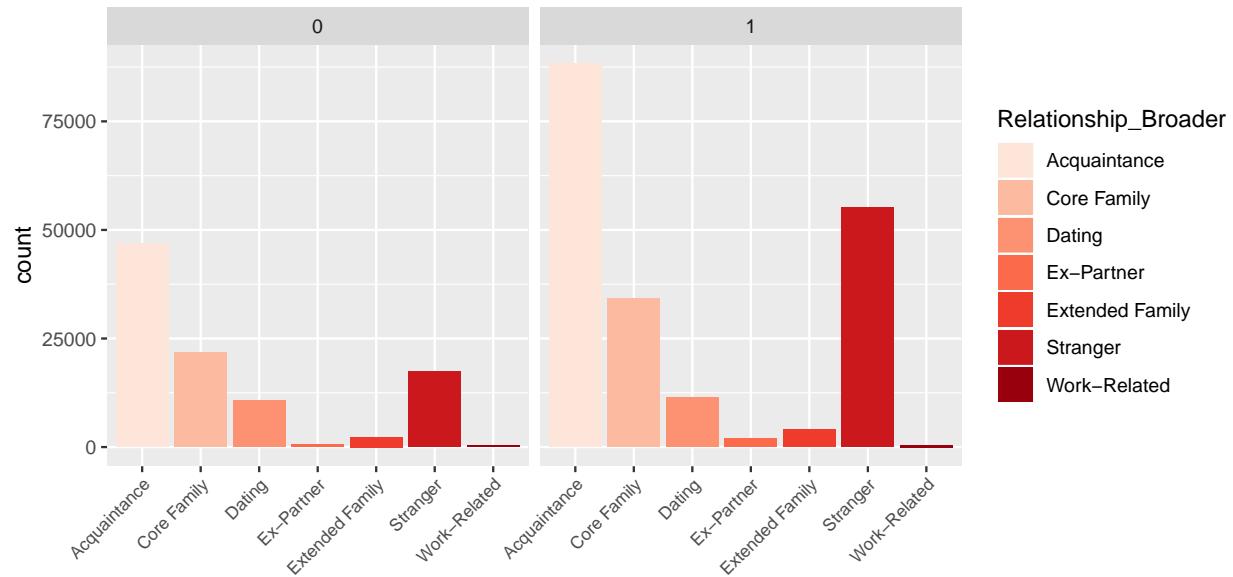
From the clusters, there does not seem to be a clear connection between Perpetrator.Age and Victim.Age. Hence, the analysis on individuals' ages is inconclusive.

6.3 Relationship between the Victim & Perpetrator



Across every region in the United States, homicides are more likely to be committed by an acquaintance, by someone the victim knows as opposed to strangers. However, the victim is more likely to be murdered by someone the victim is not very intimate with such as a friend, a neighbor, someone from work as opposed to their core family.

6.4 Weapon Preference Across Murderers



Interestingly enough, acquaintances and strangers are more likely to use firearms to commit murder whereas individuals closer to the victim (family members, ex-partners, dating) are more likely to use other methods (suffocation, poison, knife, strangulation, drowning, blunt object, explosives, etc.). Therefore, a tighter gun

control regulation could help reduce the number of homicides committed by acquaintances and strangers, which accounts to the largest fraction of homicides.

7 Conclusion

In this study, we have examined the data concerning homicide cases in the United States and uncovered a number of interesting trends and patterns.

- On the national level, the number of homicide cases have been decreasing steadily in the past three decades, although there was a significant spike in the early 1990s. Analysis of nationwide trends and some further research suggest that this spike may be due to the appearance of crack cocaine and the rise of gun ownership.
- On the state level, California has by far the highest absolute number of homicide cases, but when the population per state is accounted for the numbers of homicide / population is relatively similar across all states, with the exception of the South states having a slightly larger number.
- On the individual level, we find multiple interesting trends.
 - Homicides do not often cross racial lines.
 - Homicides are often committed by an acquaintance of the victim rather than a complete stranger.
 - Acquaintances and strangers are more likely to use firearms to commit murder whereas individuals with closer ties to the victim tend to use non-firearm methods of murder.

Homicides have been a persistent age-old problem in the United States. Even though the number of homicide cases has decreased significantly over the past three decades, the US still has a shockingly high rate of homicide cases / population for a first-world nation.

This study has uncovered some of the underlying trends that have characterized the past three decades of homicide in the country. No real conclusions on how to decrease homicide rates can be drawn as it is such a complicated issue, but the findings do tell us, possibly, where law enforcement should direct its attention towards. For example, if gun control were ever to be implemented, this report suggests the possible segments of the population who should be more heavily scrutinized in their ownership of guns based on the past tendencies of their population segments.

Ultimately, we seek to inform our readers on the underlying trends surrounding homicide in the US so that we can promote more informed discussions on this pertinent issue.