IE6600 Computation and Visualization of Data
Prof. Sivarit Sultornsanee

US Crime Analysis

Group 9

Aditya Trivedi
Amoolya Bagalkoti
Avinash Ranga
Sachin Dewangan
Yashasvi Sharma

Index

1. Introduction

In various parts of the United States, numerous forms of crimes have occurred. The purpose of our project is to provide a distinct and clear perspective on crimes that have taken place in the US by providing an interactive visualization of solved crimes in different regions. Our research intends to collect and clean crime data from numerous sources before analyzing the data to identify patterns and trends in crime rates across various states, cities, and neighborhoods.

The ultimate purpose of this project is to produce interactive dashboards and maps that allow users to explore data and spot patterns and trends using data visualization tools such as Tableau. This interactive dashboard can be used by government agencies and individuals to assess various agencies that have successfully solved crimes in various regions and how safe the neighborhood is. Different authorities can utilize this to observe patterns between different sorts of crimes and the relationship between the victim and the perpetrator.

2. Summary

Our initiative aims to provide a clear visual representation of crime rates in various regions of the United States. We hope to find patterns and trends in crime rates across multiple states, cities, and neighborhoods by collecting and purifying crime data from diverse sources.

The interactive dashboard that we are creating will make it simple for government agencies and people to identify patterns and trends in crime statistics, such as the most common crimes and the places where they occur. We will also examine the crime rate in relation to different races, ethnicities, age groups, and genders in order to discover the types of crime that occur the most frequently.

Our research will assist law enforcement organizations in making data-driven decisions and offering insights into crime patterns and trends in the United States, hence enhancing crime-solving rates. To do so, we will employ a variety of data visualization techniques and design concepts, such as bubble charts, line charts, pie charts, trend analysis, bar charts, histograms, and others.

Overall, our study will give an in-depth analysis of US crime statistics and assist law enforcement and government institutions in making data-driven decisions.

3. Data Source

Our project deals with analyzing crime data from various sources. The state_crime dataset, obtained from Kaggle, was saved in a .csv file with a size of 516 KB. The database dataset, obtained from Homicide Reports, 1980-2014 Kaggle, was also saved in a .csv file but with a much larger size of 111.8 MB. Both datasets are essential in providing a clear perspective on crimes that have occurred in different regions of the US. The state_crime dataset is useful in analyzing crime rates across various states, while the database dataset provides in-depth information about homicides reported in the US between 1980-2014. We will use data visualization tools like Tableau to create interactive dashboards and maps that allow users to explore data and spot patterns and trends.

4. Data Dictionary

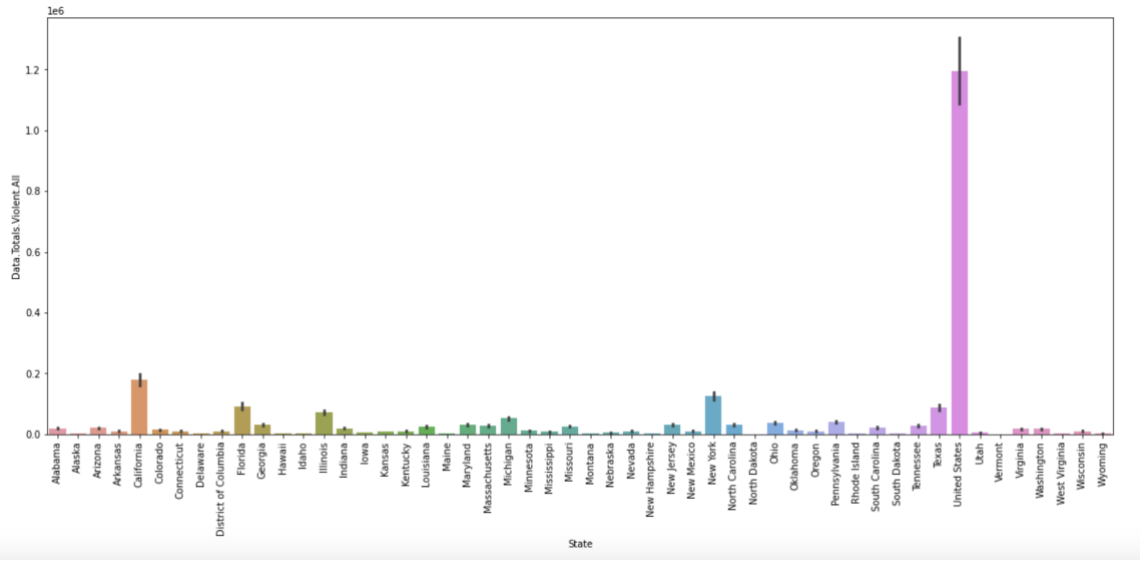
The ‘state_crime’ dataset below, contains various features such as the state name, year, population, crime rates for different types of crimes (property and violent), and total crime counts for each type. The data types in the dataset range from object, int64, to float64, depending on the feature.

Name	Description	Datatype
State	Name of the state	object
Year	Year of the data	int64
Data.Population	Population of the state	int64
Data.Rates.Property.All	Rate of all property crimes per 100,000 popula...	int64
Data.Rates.Property.Burglary	Rate of burglary crimes per 100,000 population	int64
Data.Rates.Property.Larceny	Rate of larceny crimes per 100,000 population	int64
Data.Rates.Property.Motor	Rate of motor vehicle theft crimes per 100,000...	int64
Data.Rates.Violent.All	Rate of all violent crimes per 100,000 population	int64
Data.Rates.Violent.Assault	Rate of assault crimes per 100,000 population	int64
Data.Rates.Violent.Murder	Rate of murder and nonnegligent manslaughter c...	int64
Data.Rates.Violent.Rape	Rate of rape crimes per 100,000 population	int64
Data.Rates.Violent.Robbery	Rate of robbery crimes per 100,000 population	int64
Data.Totals.Property.All	Total number of all property crimes	int64
Data.Totals.Property.Burglary	Total number of burglary crimes	int64
Data.Totals.Property.Larceny	Total number of larceny crimes	int64
Data.Totals.Property.Motor	Total number of motor vehicle theft crimes	int64
Data.Totals.Violent.All	Total number of all violent crimes	int64
Data.Totals.Violent.Assault	Total number of assault crimes	int64
Data.Totals.Violent.Murder	Total number of murder and nonnegligent mansla...	int64
Data.Totals.Violent.Rape	Total number of rape crimes	int64
Data.Totals.Violent.Robbery	Total number of robbery crimes	int64

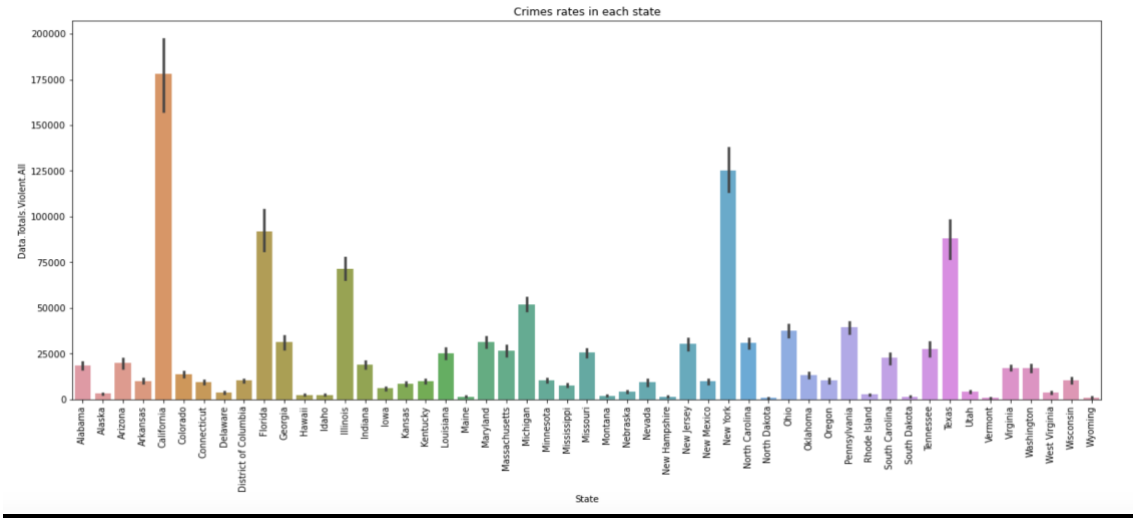
The 'database' dataset had 24 columns and 638454 rows. We discovered that the "Perpetrator Age" column had an object datatype during the preparation of the data. To eliminate this column, we substituted NAN values for empty strings and got rid of the relevant rows. Additionally, we eliminated all instances where the offender was younger than 10 years old because they were irrelevant.

Column Name	Description	Data Type
Record ID	A unique identifier for each crime incident	int64
Agency Code	The agency code of the law enforcement agency	object
Agency Name	The name of the law enforcement agency	object
Agency Type	The type of law enforcement agency	object
City	The city where the crime incident occurred	object
State	The state where the crime incident occurred	object
Year	The year in which the crime incident occurred	int64
Month	The month in which the crime incident occurred	object
Incident	The incident number within the year	int64
Crime Type	The type of crime committed	object
Crime Solved	Indicates whether the crime was solved or not	object
Victim Sex	The sex of the victim	object
Victim Age	The age of the victim	int64
Victim Race	The race of the victim	object
Victim Ethnicity	The ethnicity of the victim	object
Perpetrator Sex	The sex of the perpetrator	object
Perpetrator Age	The age of the perpetrator	int64
Perpetrator Race	The race of the perpetrator	object
Perpetrator Ethnicity	The ethnicity of the perpetrator	object
Relationship	The relationship between the victim and the perpetrator	object
Weapon	The type of weapon used in the crime	object
Victim Count	The number of victims in the incident	int64
Perpetrator Count	The number of perpetrators in the incident	int64
Record Source	The source of the record	object

5. Data collection and Data Cleaning



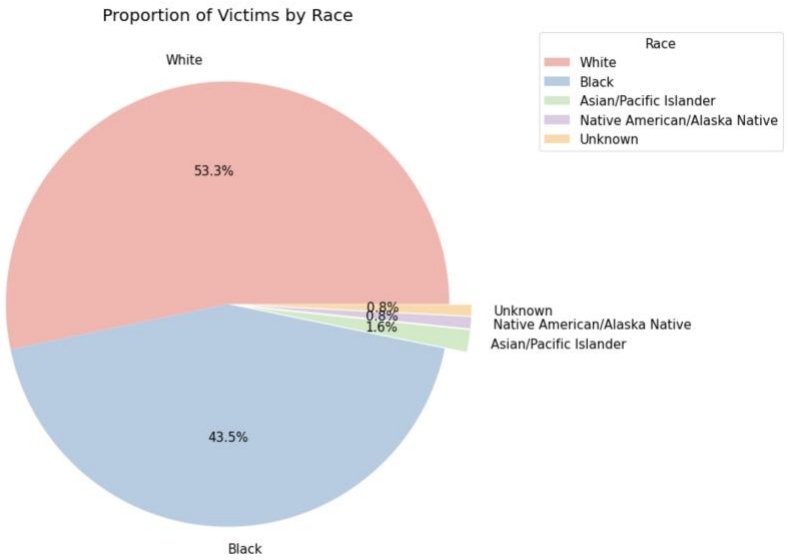
We noticed that the State field in the sate_crime dataset had a list of US states as we were conducting the analysis. Although "United States" is not a state, we noted that it was displayed as one of the data points. When compared to the other data points in the dataset, this one stood out, and subsequent research established that it was an outlier.



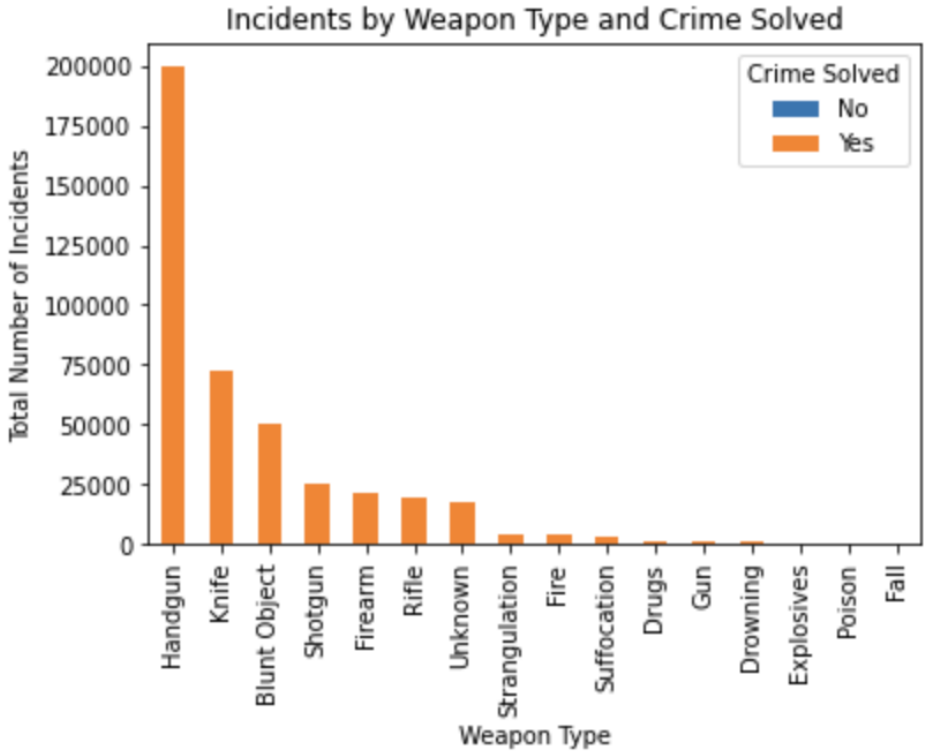
An overview of the crime rates in various states could be given by a bar chart that shows the state-by-state crime in the US. Each state might be represented by a bar of a different color, and the graph may display the total number of crimes committed in each state. The graph might also show which states have the highest and lowest rates of crime, giving important information about the elements that influence crime in various areas.

6. Data Visualization

Our examination of US crime statistics tries to reveal trends and patterns in crime across numerous states, cities, and neighborhoods. We have developed interactive dashboards and visualizations that let consumers explore data and readily identify patterns using data visualization tools like Tableau. In the section that follows, we showcase a few of the visualizations that offer a clear understanding of US crime patterns and rates.

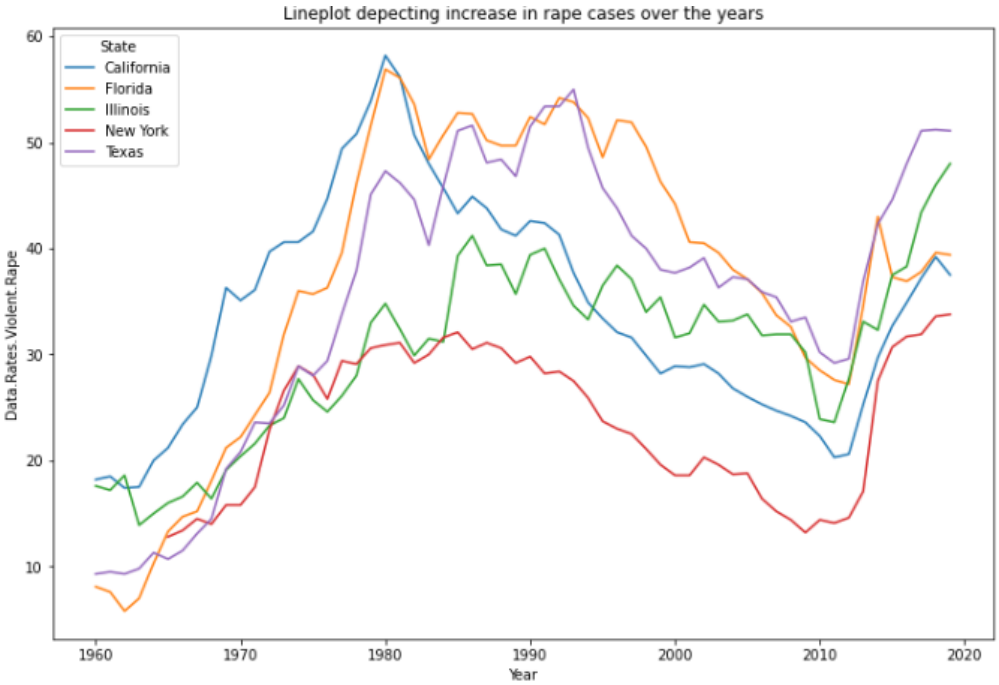


The pie chart we used in our analysis of the US crime data offers a useful breakdown of the racial groupings that were impacted by crimes in the dataset. According to the graph, more than 50% of the victims were White, followed by Black victims, who made up roughly 40% of the total. The graph also emphasizes the lesser proportion of victims from other racial groups, including Asian, Unknown race, and Native American, who combined account for less than 10% of the overall casualties. The pie chart makes it easier to see how crimes affect various racial groups, which is helpful when coming up with plans to lower crime rates and increase safety in the impacted areas.



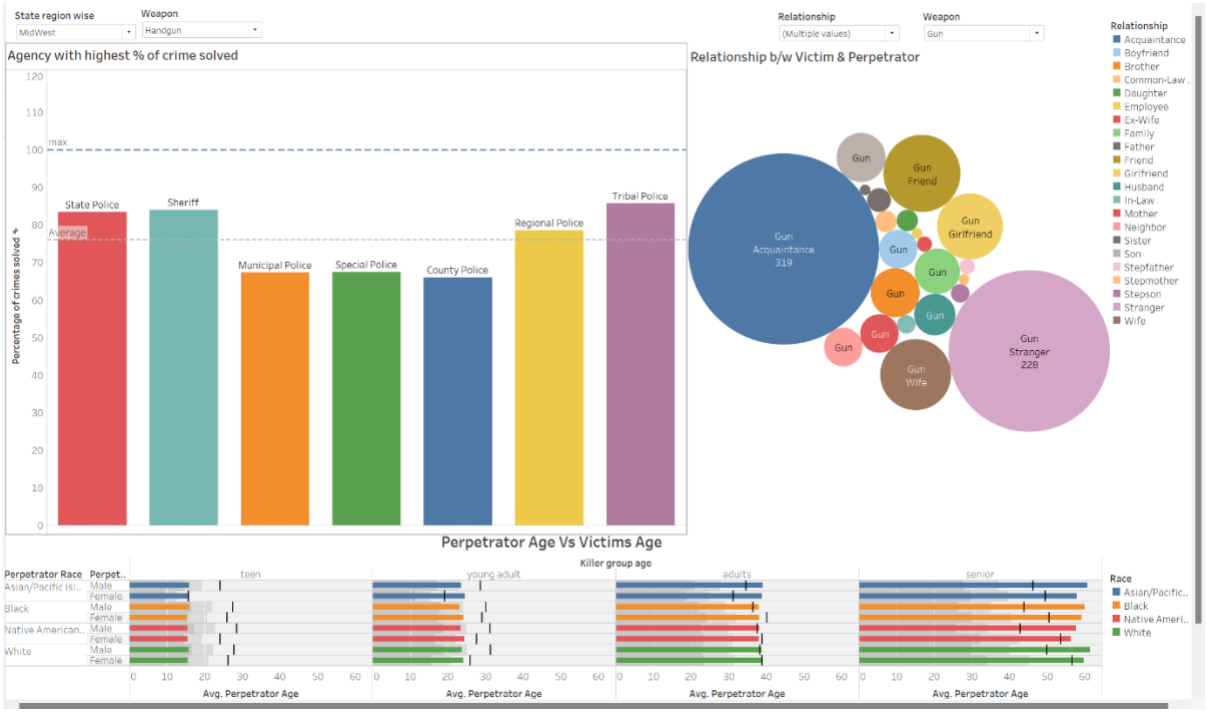
The stacked bar graph graphically illustrates the association between the types of weapons used in events and whether the crime was solved. The graph is split into two parts, one showing case where the crime was solved and the other showing incidents where it was not. The chart's y-axis lists the total number of events for each weapon type, while the x-axis lists the various weapon kinds that were employed in those incidents. The percentage of occurrences resolved for each weapon type is shown by the stacked bars. This makes it simple to compare the efficacy of law enforcement in resolving crimes involving various types of weapons.

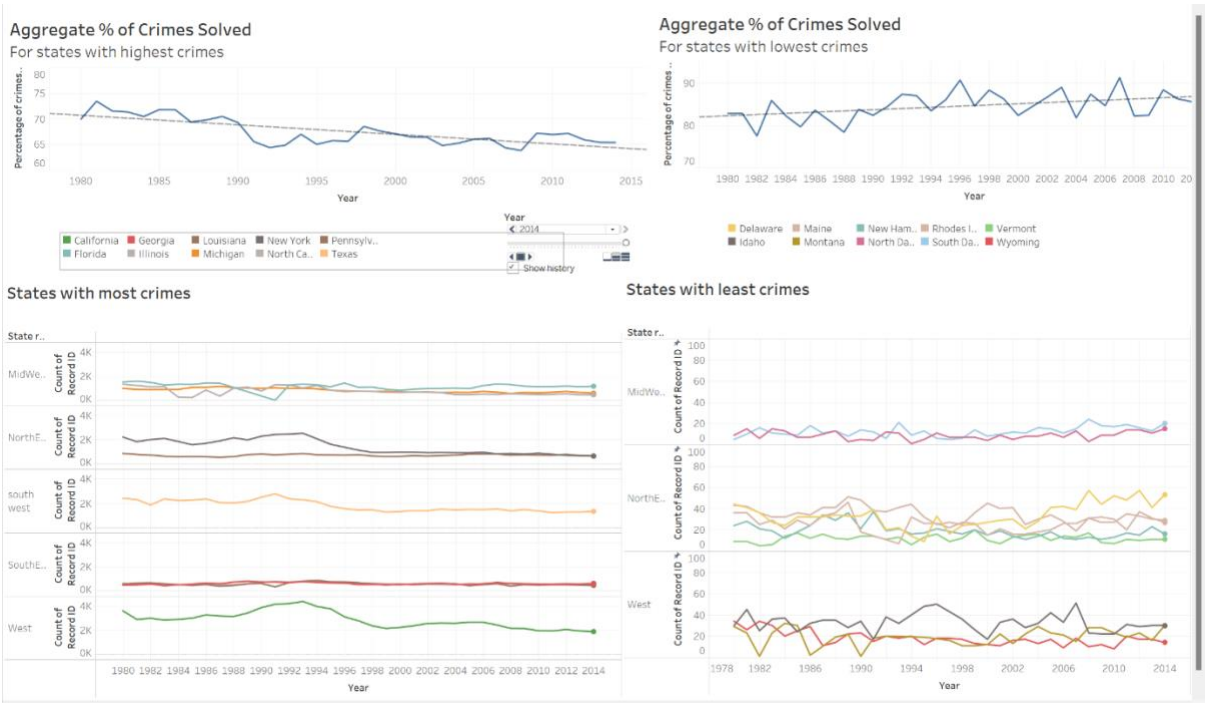
The orange color in the stacked bar chart shows crimes that were solved, while the blue tint represents crimes that were not solved. Surprisingly, the blue hue in the chart is scarcely visible, indicating that over 99% of the events were resolved and closed. This is a good outcome that indicates that law enforcement authorities in the United States have done an excellent job in solving crimes.



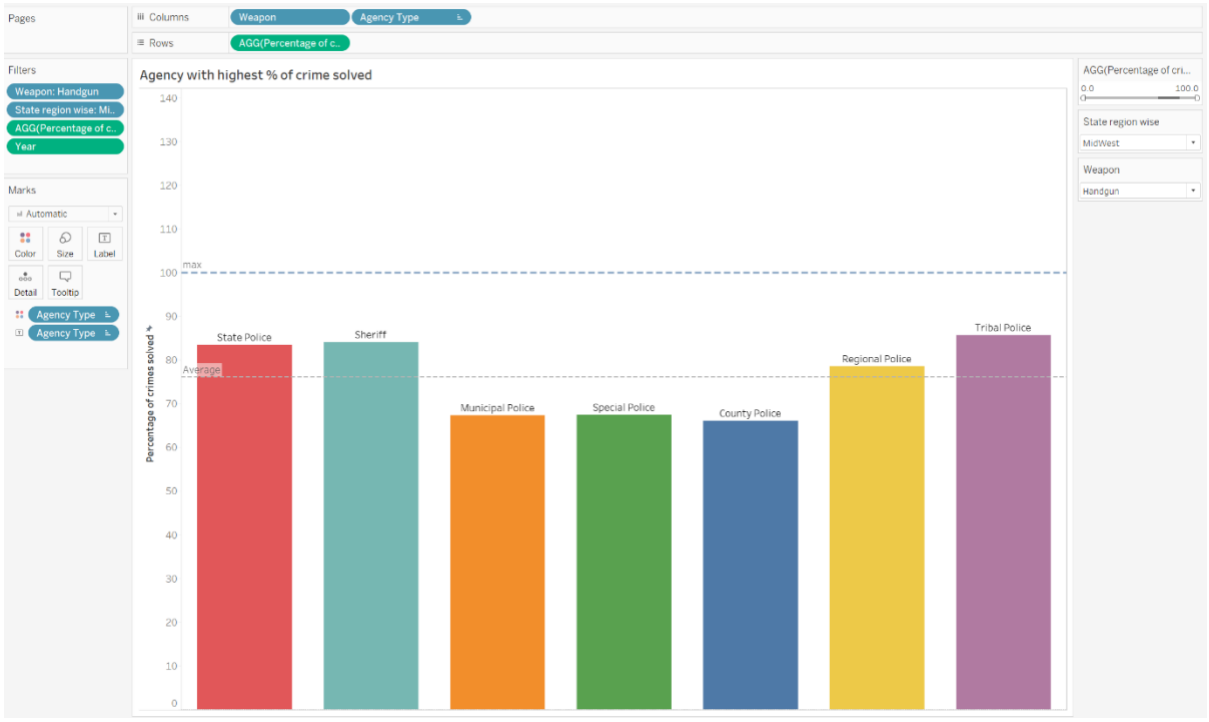
Our line graph depicts the frequency of reported rape cases in California, Florida, Illinois, New York, and Texas between 1960 and 2020. The graph shows a considerable increase in rape cases during the 1970s and the 1980s, followed by a gradual drop until 2010. However, the graph shows a rebound of rape cases in these states after 2010, showing a reversal in the downward trend.

Visualization and Dashboard in Tableau:



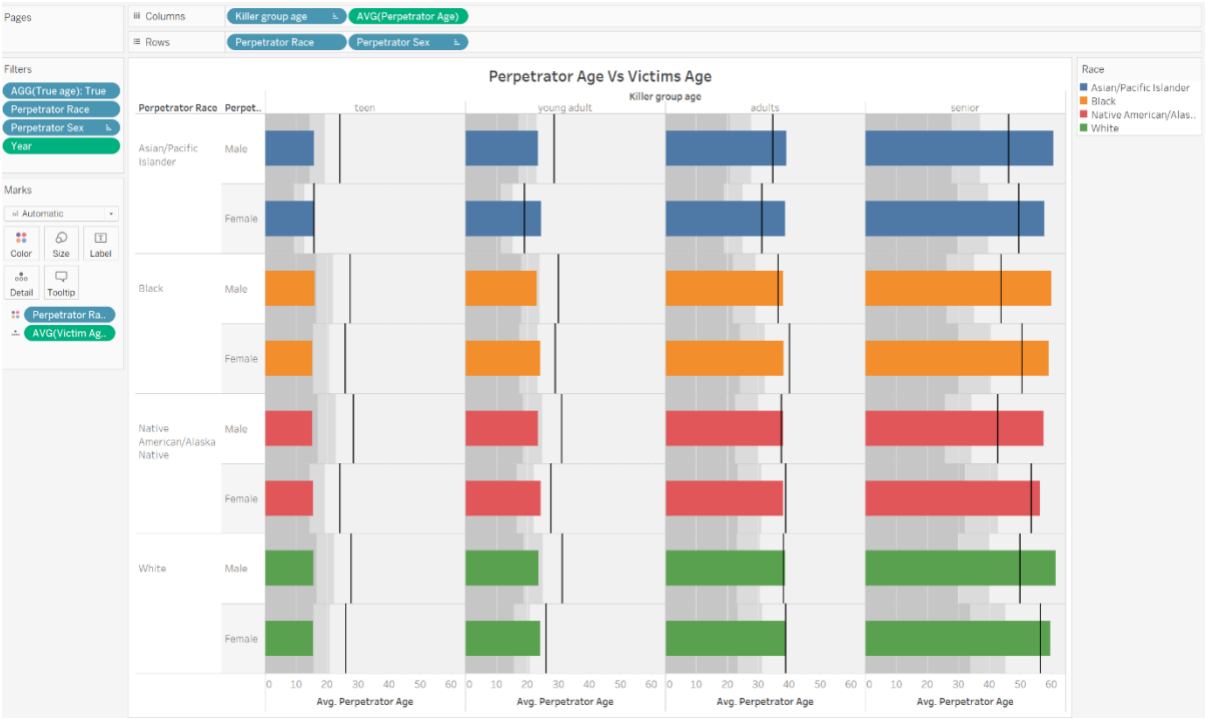


Bar Graph:



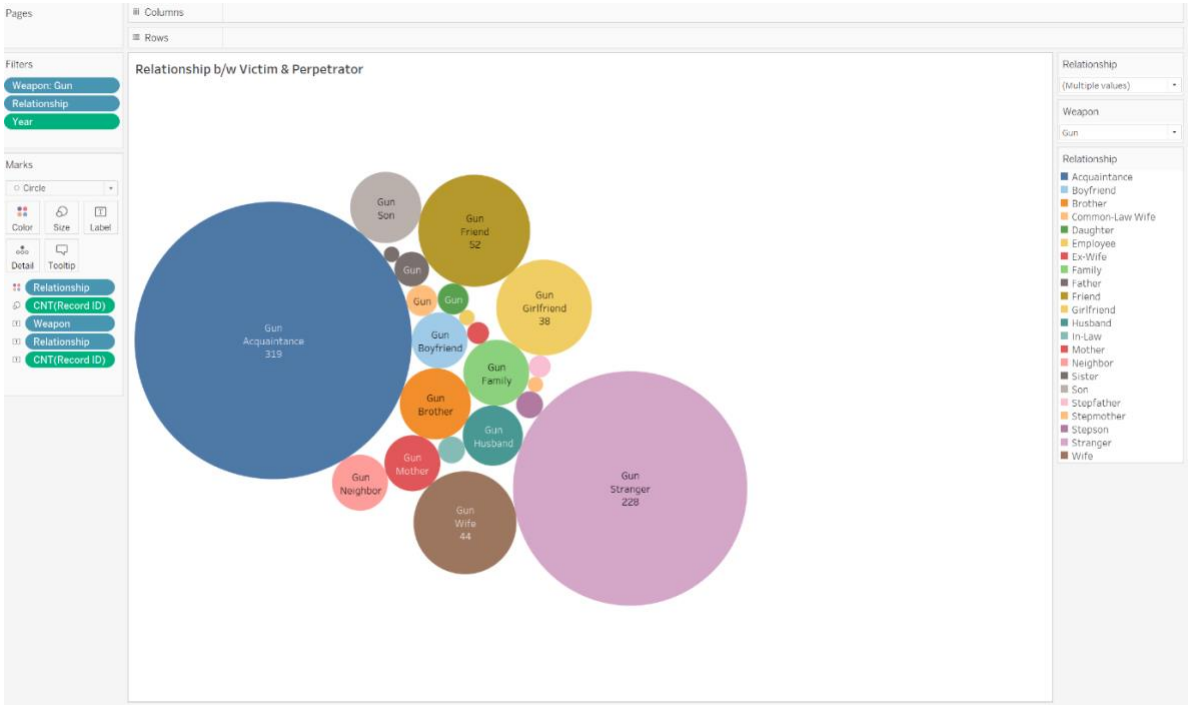
The goal of this visualization is to depict crimes solved by various authorities based on the weapons used in the crime. Because the dataset only contains states, we are constructing groups (east, west, central, north, and south) for the state variable to make it easier for the user to infer data dynamically in different parts of the United States. Creating groups also aids in the reduction of clutter in the final dashboard. The Y-axis represents the type of weapon and agency. On the X-axis, the percentage of crimes solved (calculated field). To make the visualization interactive, the state is utilized in filters, and the Agency type is used on color in the marks card. The percentage of crimes solved is shown by the length of the bar. The greatest way to perceive agency type, which is a categorical attribute, is through color hue.

Bullet Graph:



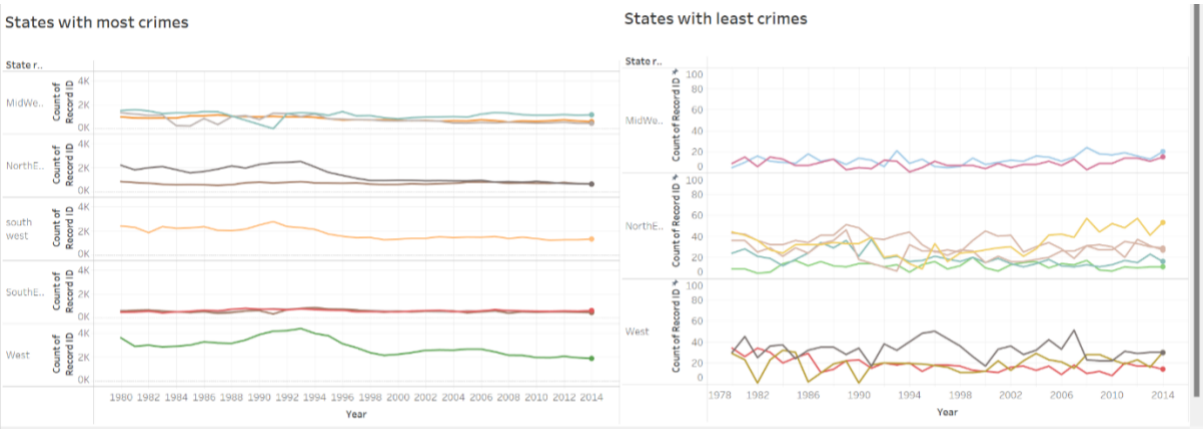
The goal of this visualization is to determine the age difference between the victim and the offender depending on their gender and race. After constructing a few computed fields, we determine the average age of the offenders and categorize them into four age categories (youth aged 9 to 18, young adults aged 18 to 35, adults aged 35 to 60, and seniors aged 60 to 100). The length of the bars in the graph shows the average age of groups from a certain race or gender based on the age of the perpetrators, and the black line represents the average age of the victims killed by that category of the perpetrator. This allows us to determine whether a race or gender's older or younger age groups commit crimes against one another. Here, the y-axis shows the perpetrator's race and gender while the x-axis shows the average perpetrator age divided into 4 age groups. Race and sex of unknown perpetrators are removed using filters. The color of the markings card indicates the race of the perpetrator, and the age of the average victim is displayed in the marks card's description. This makes the chart more appealing to the eye.

Packed Bubble Chart:



This visualization tries to examine the perpetrator's weaponry and their relationship with the victim. We will utilize a bubble chart for this visualization, with the number of crimes committed obtained by counting the record ids. The size of the bubble represents this quantitative value. The bubble's colors are utilized to differentiate between the weapons of choice, which is a qualitative property. Because three-dimensional properties can be employed at the same time, bubble charts provide excellent visualizations. The weapon and relationship traits are both employed as filters, allowing us to see how a perpetrator's relationship can be used to locate the weapon of crime.

Line Graph:



This visualization helps us analyze crime trends since 1976. Group-(north, east, west, center, south) generated from the state attribute to visualize which part of The United States has more crime in a 30-year period. Line charts are used to compare changes in the same period for more than one group. Furthermore, since this is historical data Line chart is the best to use. The Year property is on the Y axis. Status attribute and ID number enabled X-axis. ID attribute, which is quantitative data, represented by location. Add more crime than bad points. We also distinguish each state according to color. We can observe the Gestalt principles of visual perception in the picture. If there is continuity in the data, we can see the forming principle of continuity and connection. From this is time series data, to show the annual crime rate over the years the year attribute would be added to the pages making the dashboard come to life.

6.Limitation and Future Scope:

1.Data quality and consistency: The quality and consistency of the data are one of the main constraints of crime data analysis. Law enforcement agencies' reporting procedures determine how accurate the data is, and inconsistent data gathering procedures might make it challenging to infer useful conclusions from the data.

2.Limited scope: Since unreported or undetected offenses are often not included in crime statistics, there may be an incomplete picture of crime trends.

3.Bias: Due to variations in reporting procedures and enforcement procedures across jurisdictions, the statistics may be biased.

4.Privacy issues: Crime data may include personal information that is sensitive, which can cause privacy issues.

5.Lack of context: Without additional contextual information, such as socioeconomic characteristics, demographic statistics, and other pertinent elements, crime data alone cannot provide a complete picture of crime trends.

In the future, it may be possible to evaluate crime data more skillfully utilizing cutting-edge technology like machine learning and artificial intelligence. These technologies could be able to spot patterns and trends in crime data that conventional approaches might miss. Additionally, combining information from several sources can offer a more thorough knowledge of crime trends. Additionally, there is a chance to use crime data to create tactics for crime prevention and intervention that are more successful.