**CSCI 55200** - **Data Visualization**

**Term Project Report**

**Visualization of Crimes in the USA using D3**

Nirmal Kumar Raveendranath

Bhavani Prasad Rao Ejanthkar

**Summary**

This Report contains 9 visualizations with dashboard inclusive. It also contains snapshots of various visualizations and insights describing each visualization.

**Table of Contents**

Summary i

Abstract & Background 4

Description 4

Dataset 4

Visualization Techniques 6

1. Scatter Plot matrix 6
2. Parallel Coordinates 6
3. Geo Maps 7
4. Heat Maps 7
5. Stacked Bar Chart 7
6. Line Chart 7
7. Pie Chart 7
8. Dashboard 7

Visualization Snapshots with Insights

1. Parallel Coordinates 8

1. Scatterplot Matrices 9

1. US heat map of Crimes in 2015 10

1. Crimes per population (heat map in bubbles) in each city for year 2015 11
2. City with total number of Crimes for year 2015 12
3. Stacked bar chart with crime type and each state for year 2015 13
4. Interactive Bar chart 14
5. Interactive Pie Chart for year 2000 15
6. A Collective Dashboard 16

Conclusion 18

References 18

**Abstract:**

In this project, we did a visualization on Crime dataset that was taken from various sources of FBI. This dataset consists of the factors that influencing the crime rate in the states, and that will give the correlations and trends in the data. For this we used 7 visualization techniques that was learnt in class that will give the insights for the given data, they are Multi-dimensional Visualization techniques like Scatterplot Matrices, Parallel Coordinates, Geo maps, Heat maps, Pie Chart, Stacked Bar Chart, Line Chart, Dashboard.

We are finding the Correlations, Trends, and Factors that are contributing to the Crime Rate in USA. We have used two crime datasets that give the insights for the crimes in the specific type of crime and city and other used for the overall clear big picture for the Crime Rate in States and influence of attributes like Population, Socio-economic indicators(GDP), social attitudes (Literacy Rate, Unemployment Rate) on the crimes in USA.

**Background:**

Crimes are the biggest concern in USA. The crimes rate varies from state to state and from city to city in a geographic location this is clear in the FBI’s data. Some states have the higher crime rate are Alaska, Nevada, New Mexico, and Tennessee. Whereas, States that have the lowest crime rate are Maine, New Hampshire, Vermont had the nearly 1/3rd of the crime rate of the states with highest crime rate. *Example:* Highest Crime Rate States have 600 violent crimes per 100,000 residents and the Lowest Crime Rate States have 200 violent crimes per 100,000 residents.

It is clear from the FBI’s data and statistics that the crime rates in states are influenced by various factors like population density, and socio-economic indicators (GDP), and literacy rate. It can be clearly noted that the states with the lowest crime rate are all from the Northeast part of the USA (Maine, New Hampshire, and Vermont).

**Description:**

**Tools Used:** D3.js, DC.js, Crossfilter.js, HTML, CSS, SVG

**Dataset:**

There are two datasets that have been used in this project. One dataset corresponds to the crimes in each year and another how crimes is related to other attributes like GDP, population likewise. On First dataset, we visualize the crimes according to years and see the trend whether the crimes have been decreased from its previous years or not.

The first dataset has attributes like year, types of crimes and how many crimes in each type of crimes. The crimes are associated with each state is shown. We see the trend from the year (1996-2016). Dimensionality of dataset is 20(years)\*51(states)\*8(Types of crimes)

The features in the dataset are

* Year
* Violent\_crime
* Rape Legacy
* Roberry
* Aggravated Assault
* Property Crime
* Burglary
* Larceny
* Motor vehicle Theft

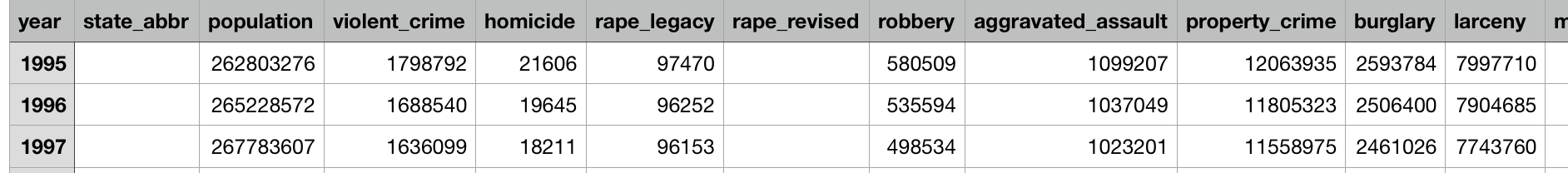
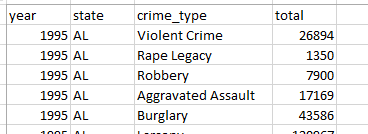


Fig 1 (Dataset of Crimes in each states)

A Formulated dataset for Crime Dashboard

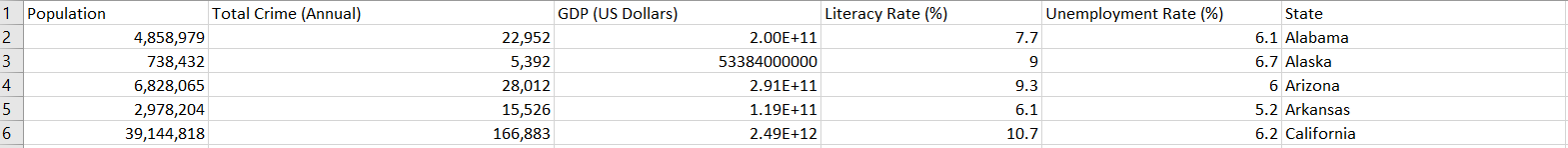


The Second Dataset is constructed by taking features (attributes) from various sources for the year 2015. Then the data is normalized, cleaned, and preprocessed such that the data across the dataset is consistent. The dataset is constructed such that the attributes taken are related to the Crimes in USA. The dimensionality of dataset is 51 x 6, where there are 51 rows that corresponds to the 51 states of USA and 6 features(attributes).

The features in the dataset are

* Population
* Total Crime (Annual)
* GDP (US Dollars)
* Literacy Rate (%)
* Unemployment Rate (%)
* State

The dataset is in csv format.

****

***fig*: crimes in the USA dataset**

**Visualization Techniques:**

**Scatterplot Matrices:**

This is a Multi-dimensional visualization method is used to show the correlation between pair of variables. If there are n variables, then there will be n-choose-2 pair of variables in a scatterplot. This Scatterplots are organized into matrix will form a scatterplot matrix that shows the pairwise correlation for all n variables.

In this scatterplot matrix visualization method, we have taken all the features (attributes) in the dataset and found the correlations among the pair of variables. Then we have considered the **State** attribute as the data point for the method.

In our dataset, we have 5 attributes. We have visualized the scatterplot matrix that shows the pairwise correlation between the variables. So, we have observed that the 10 correlations were formed and the dataset all the attributes are **positively** correlated. There exists a **Strong** correlation relationship between **Population vs Total Crime, and Total Crime vs GDP.**

**Parallel Coordinates:**

This is a Multi-dimensional visualization technique is used for the plotting multivariate, numeric data. This visualization is best for comparing the many variables (attributes) together and getting correlations between them. In this technique, all the variables are given axis and this axis are arranged in parallel to each other. Different axis has different scale and these values are normalized to keep all scales uniform and each value is plotted as the series of lines connecting between the axes. This connected line can be used to find the patterns or correlations between variables (attributes).

Since, our data is not data-dense so haven’t used interactivity technique called **Brunshing.**

In our dataset, we have observed that most of the series of lines are connected with population under 15 Million. There are patterns or correlations with the variables (attributes) **Population**, **Total Crime**, **GDP**, **and Literacy Rate**.

The best pattern that describes the crime are 90% of crime in **State** that has the **Population < 15 Million, GDP< 800 Billion Dollars.**

**Geo Maps:**

Visualization of crimes by geo maps gives more insight on how and which states have more crimes and also it can help in giving a real picture of the location of crimes. Its main attributes are states names and with the help of state name, geometric structures are drawn to give the shape of the state, the Latitudes and Longitudes coordinates of the states makes a polygon in the map. With the help of the Polygons, the actual US map like is constructed. This framework is constructed in d3.js we use this framework to show crimes in US map.

We use states, year and type of crimes data to show different visualization of the geo maps. Each visualization gives an insight about crime data without the need to look through the csv data. With this visualization, we can come to know which state highest crime rate has, which state has lowest crime rate and also according to data. It also gives information on cities crime data.

In our case, we have total crimes against each state and represent crime statistics on each State wise of US. This gives information on Crime rate on each state and also total crimes on each year, on basis of years.

**Heat Maps:**

Visualization of crimes by heat maps gives an additional advantage on geo maps, it helps in giving an overall picture of US crimes from all states data. Heat maps determine how crimes are happening in each state and also in each city.

This heat map helps in showing how crimes per population are affected in each state and crimes per 100k per population.

**Line Chart:**

We have the attributes like No of crimes vs population, with this line graph is constructed. Line graph gives information on trends like in which years **crimes has deaccelerated or accelerated verses Population, years**

**Pie Chart:**

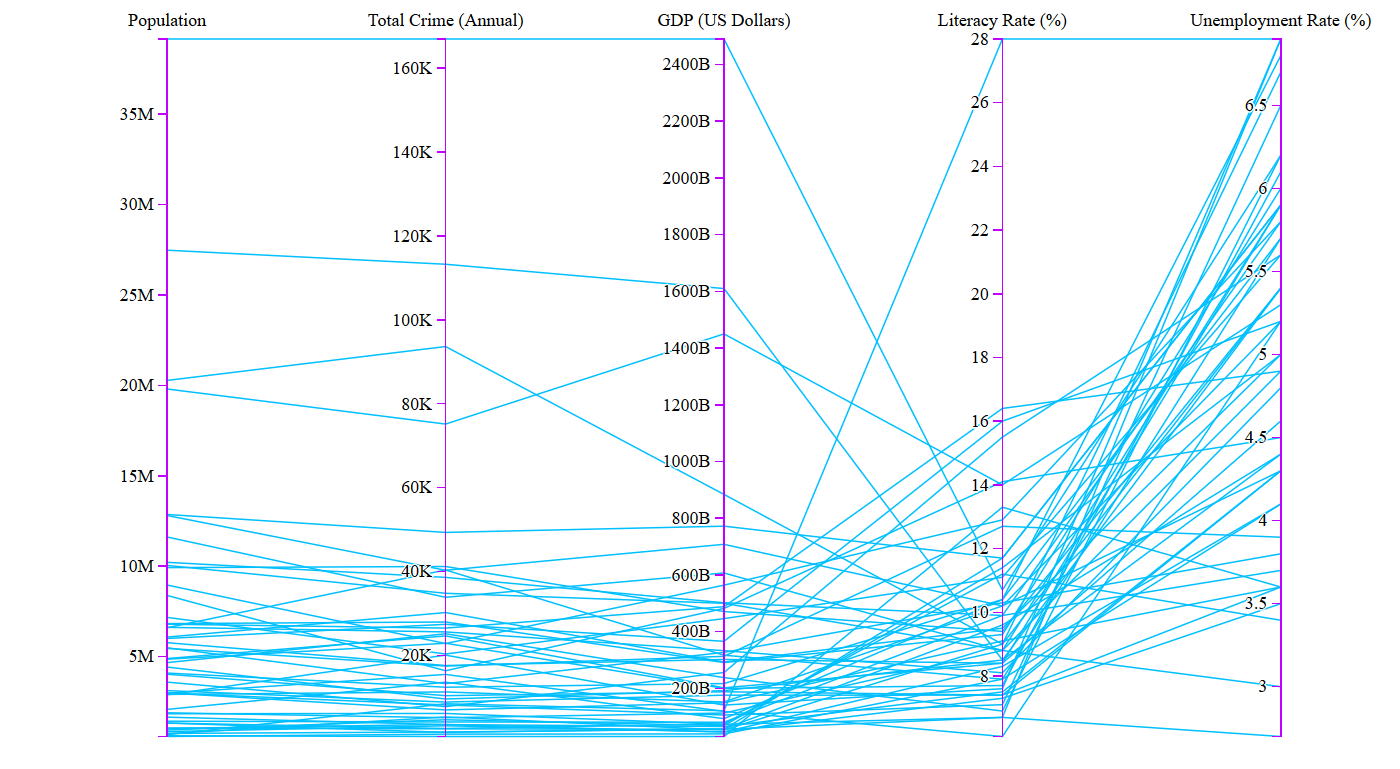
This chart gives a percentage of crimes in each category. It helps in visualizing which **type of crime has greater weightage against all type of crimes**.

**Dashboard:**

This gives overall view of the Crimes in each state in US with each type of crimes total and overall total. This can be got using Pie chart, Line chart and geo maps. Years vs Crimes and each type of crime total is given in Line chart. Pie chart is another way of giving visualization of the crimes in total. Each state can be clicked to get in detail of crimes in each of the charts and also

**Visualization Snapshots**

**Parallel Coordinates**



**Insights:**

We have noticed that the variables have trends in the crimes. That is, more than 90% of **States** have the crimes that have the **Population < 15 Million** and **GDP < 800 Billion** and **Literacy Rate < 16%**.

***Unemployment Rate*** is not a good indicator for the Crimes this concludes in our visualization that there no trends associated with the all other variables(attributes) in the dataset.

**Scatterplot Matrices:**

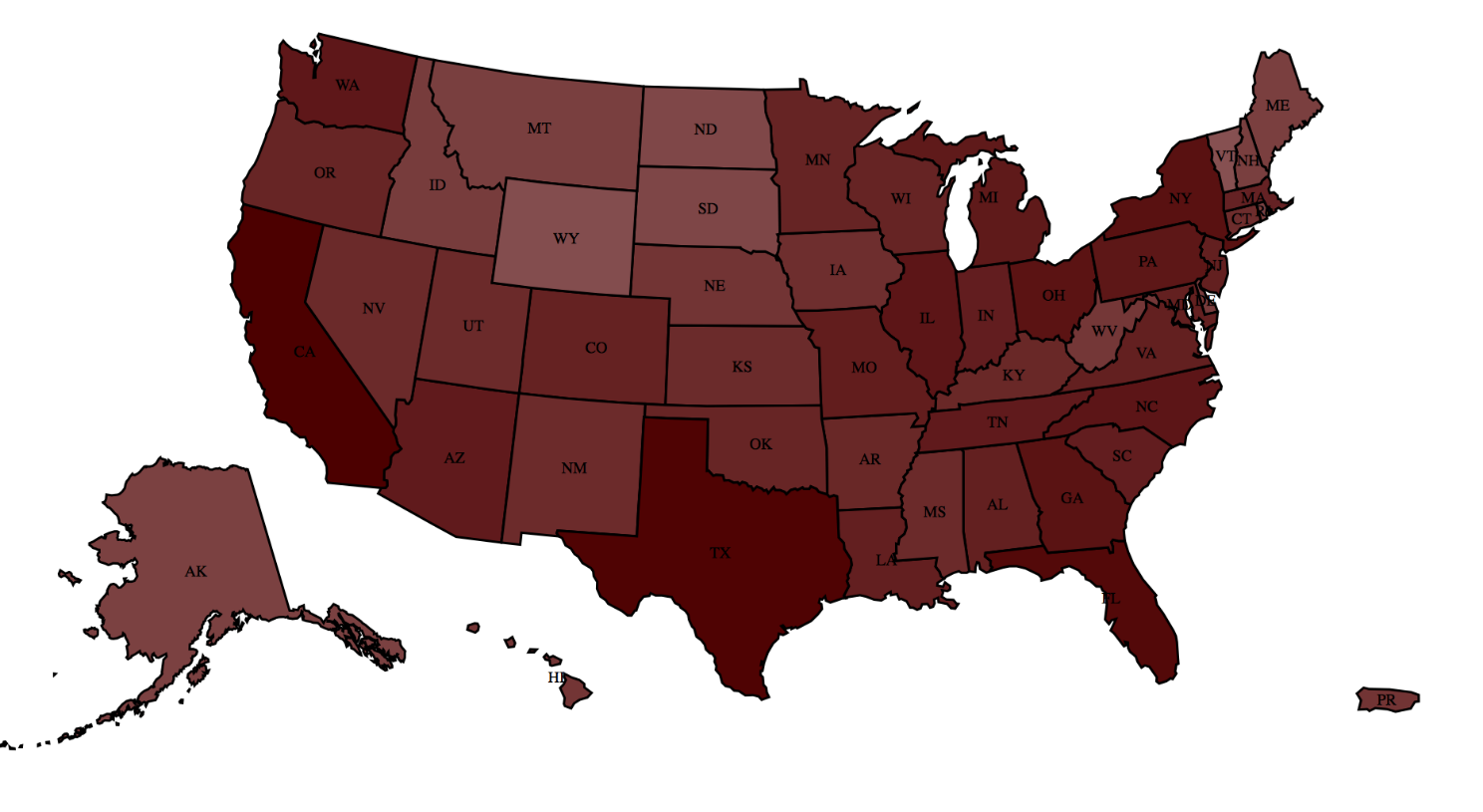
**A close up of text on a white background

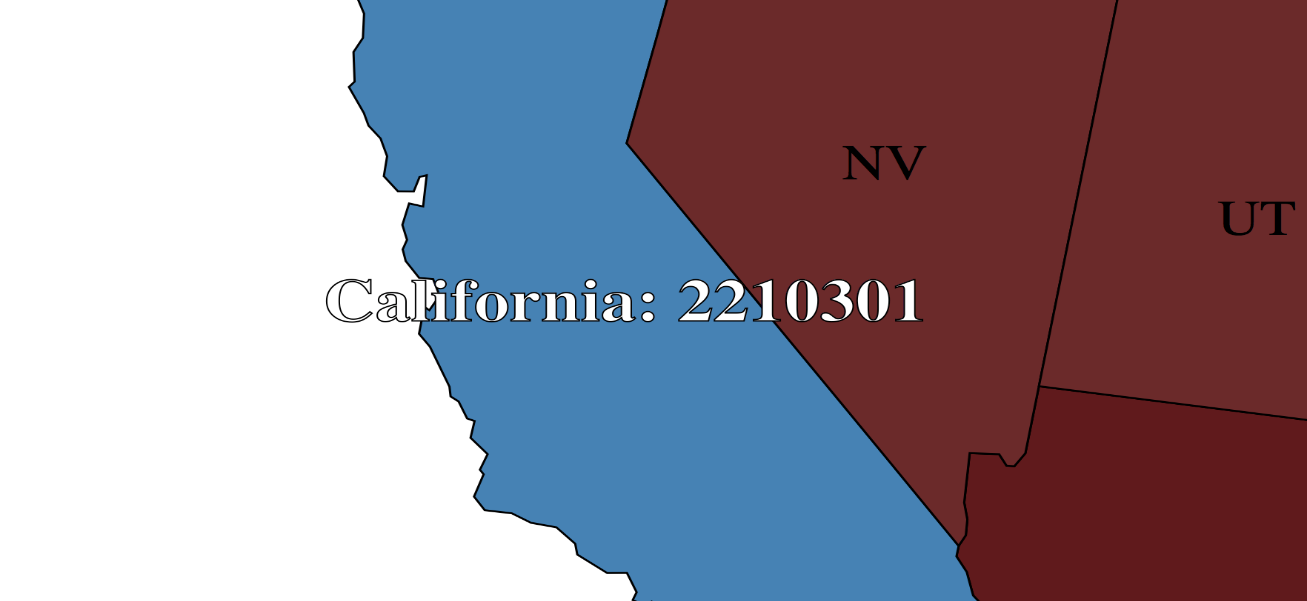
Description generated with high confidence**

**Insights:**

We have observed that all features are ***positively*** correlated and there exists a ***Strong*** correlation between **Population vs Total Crime** Scatterplot i.e., the state that as higher the population has the higher the crime rate. And also, **Total Crime vs GDP has a Strong correlation** i.e.,the state with the higher GDP has the higher crime rate**.**

**US heat map of Crimes in 2015**



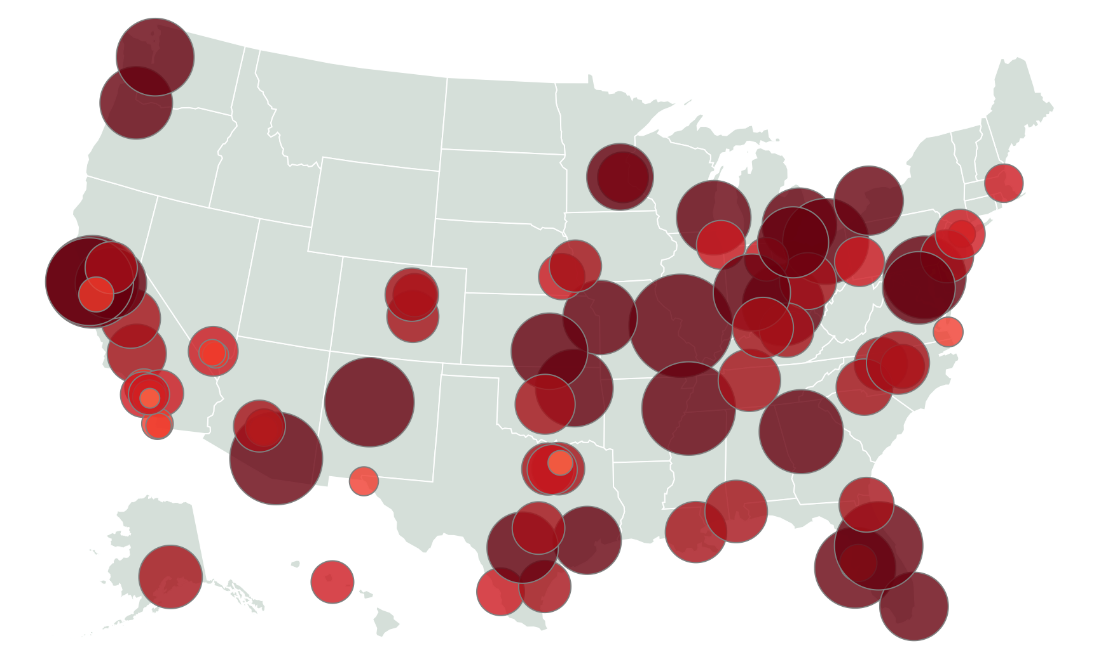


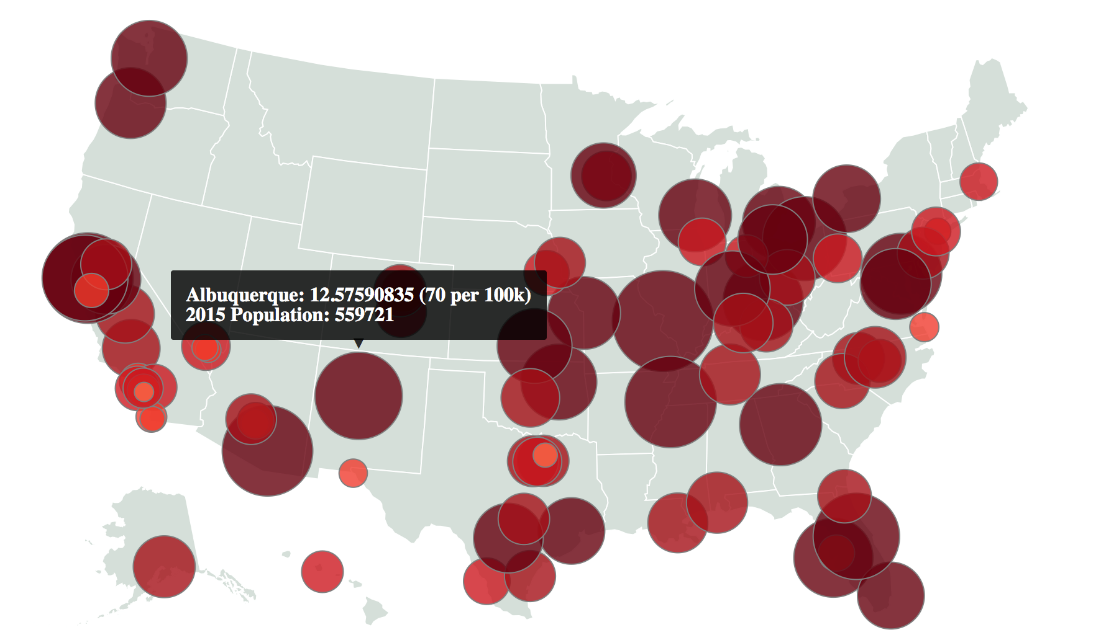
This is a heat map determine crimes in US states with total number of crimes in each state as the parameter.

**Insights:**

We can see that **Texas** and **California** is having largest number of crimes in total for the year 2015. Next other state is **Ohio.** By determining the **thickness** of the color of each state, it gives the **ranking** of each states in crimes in total. This heat map is interactive. if a **state is clicked**, then we will get the **number of crimes in total**.

**Crimes per population (heat map in bubbles) in each city for 2015**

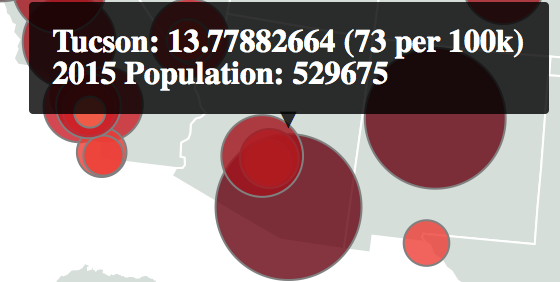




This is a heat map determine crimes in US states with total number of crimes in each city per population as the parameter.

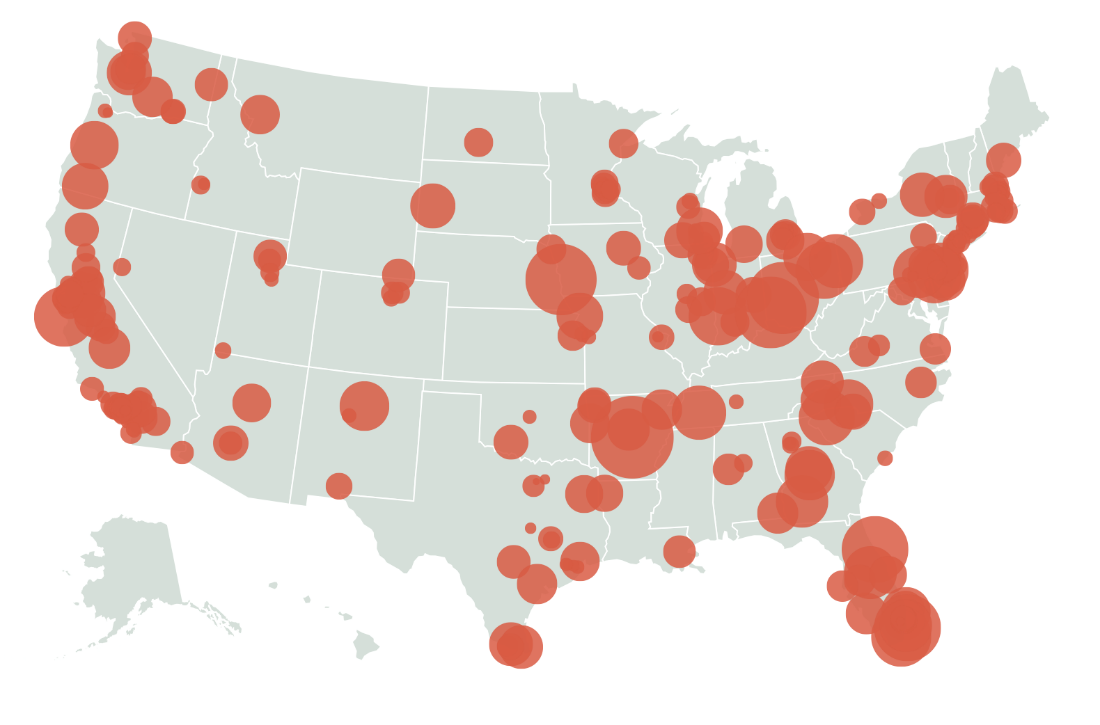
**Insights:**

A total of **top 100 cities in Crime** list is taken and shown in the geo map. Each **bubble** represents **ranking** of city in crime depending on the population, we have used **crime per population ratio** in each city to determine the size of the bubble. When the bubble is hovered, it shows a tool tip showing the ratio like this



Total Population/100,000 gives a ratio with total number of crimes in each city

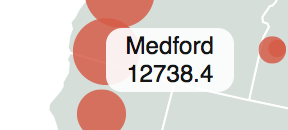
**City with total number of Crimes for year 2015**



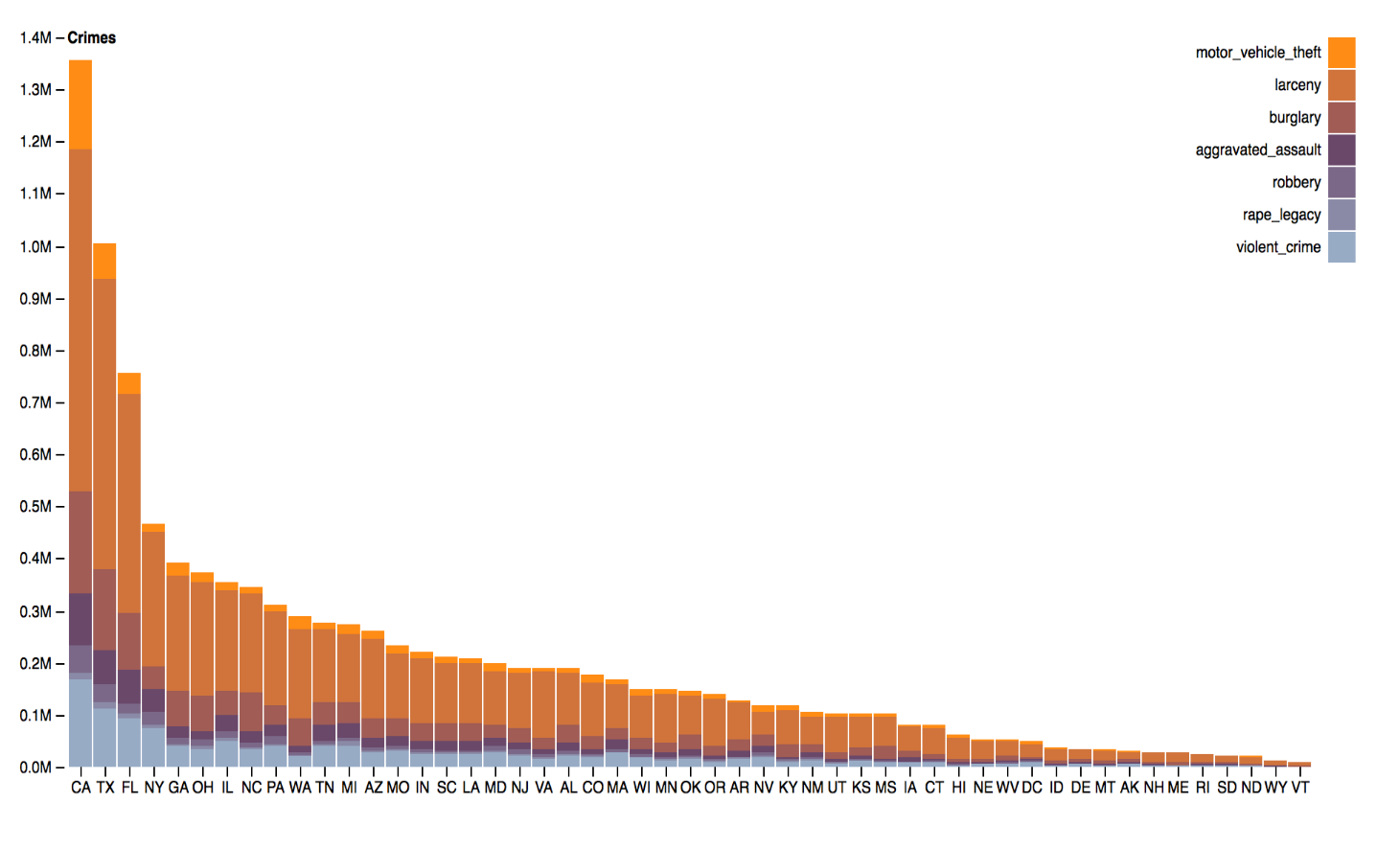
This geo map gives ranking of Crimes in each city

**Insights:**

A total of **top 100 cities in Crime** list is taken and shown in the geo map. Each **bubble** represents **ranking** of city in crime, we have used **total number of crimes in each city determine the size of the bubble**. When the bubble is hovered, it shows a tool tip showing total crimes in each state



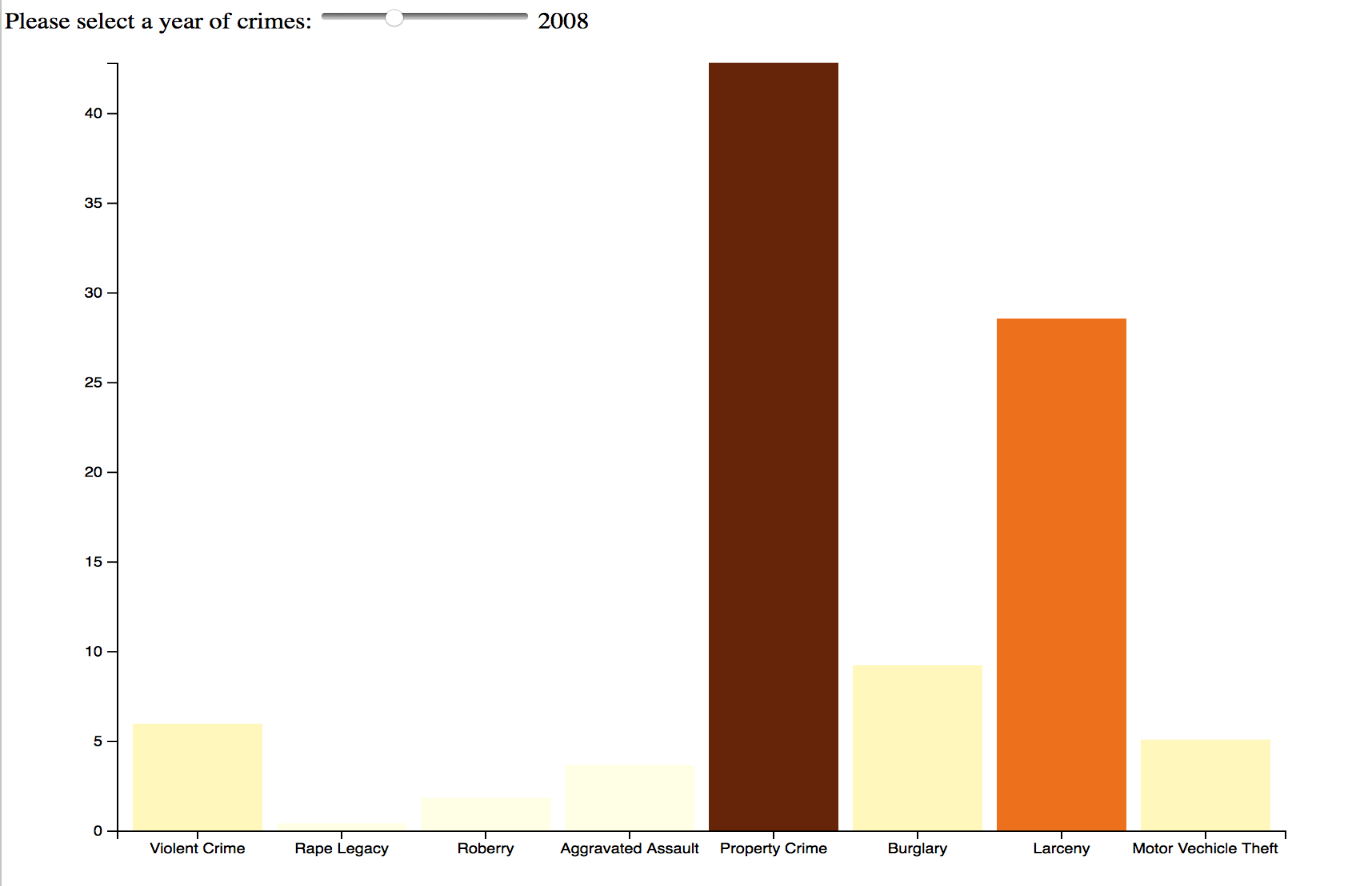
**Stacked bar chart with crime type and each state in 2015**

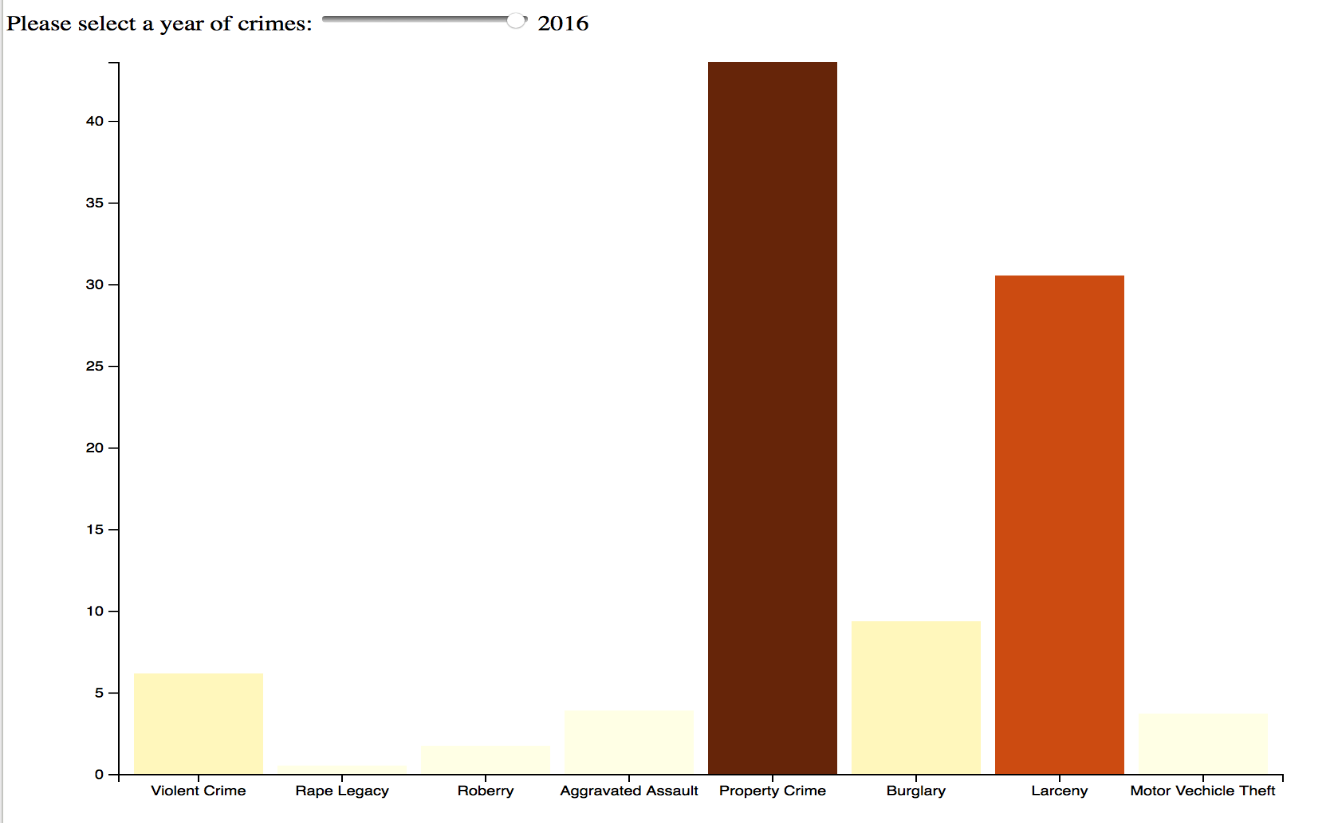


**Insights:**

Stacked bar chart shows total no crimes in each crime category. From this stacked bar chart, it clearly shows **California and Texas** having highest number of crimes. California is having crimes between **1.3 to 1.4 million** in 2015. **Larceny** category has highest level of crime rate in CA.

**Interactive Bar chart**

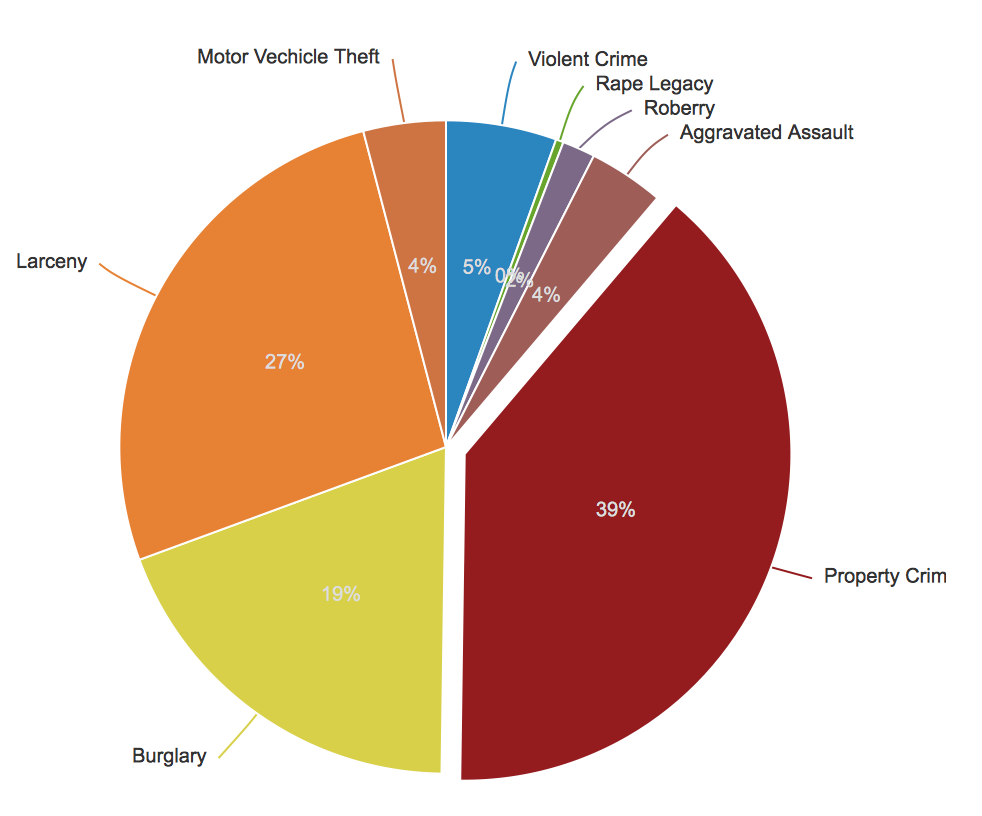




**Insights:**

This interactive bar chart shows the **difference** in crimes from year 2004 to 2016. **Difference between years** is calculated and the difference is shown in graph as **percentages.** From this graph you can see how larceny has been **increased from 25 % in 2008 to 30% in 2016** and also **Motor Vehicle theft has been reduced from 5% to 2.5%.**

**Interactive Pie Chart for year 2000**

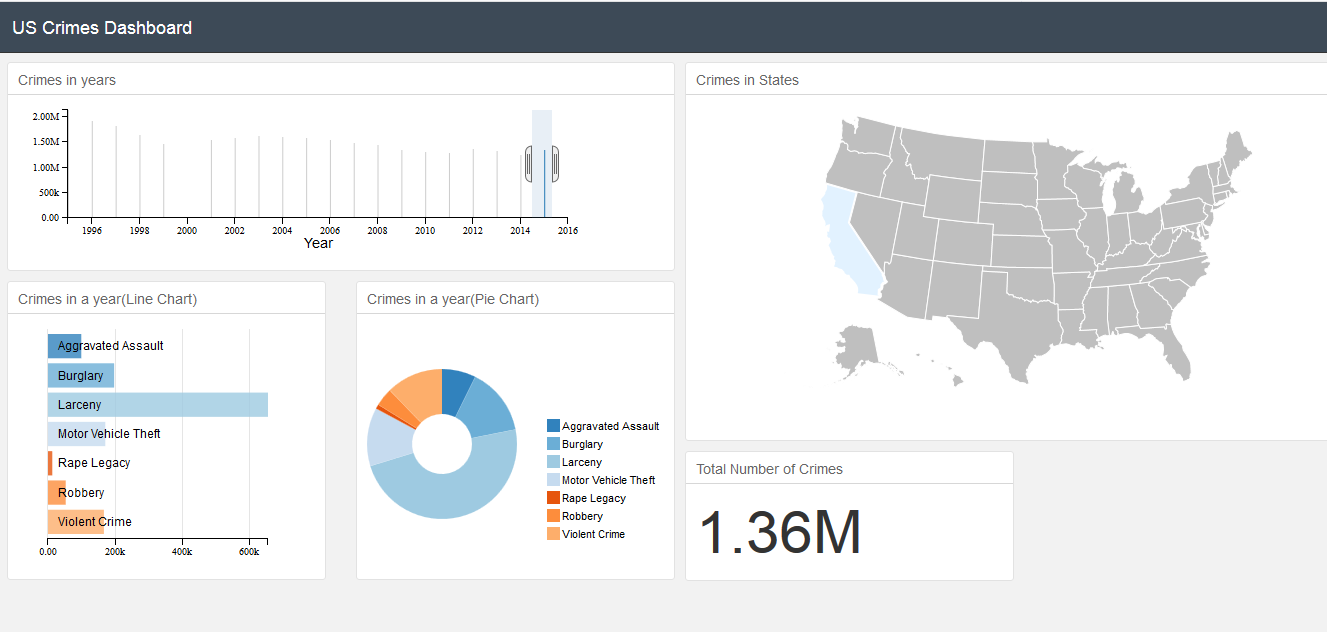


**Insights:**

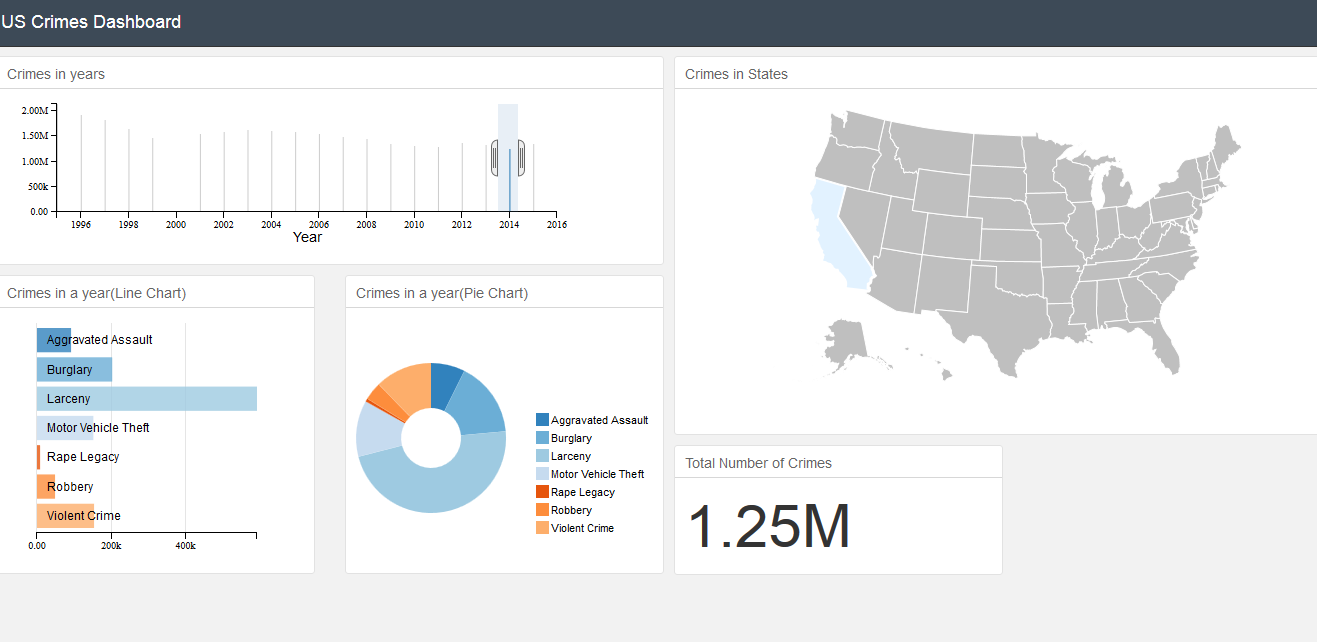
This pie chart gives Crimes in % for the year 2010 by collecting all states crimes as total. In total the largest share goes to **Larceny.**

**Collective Dashboard**

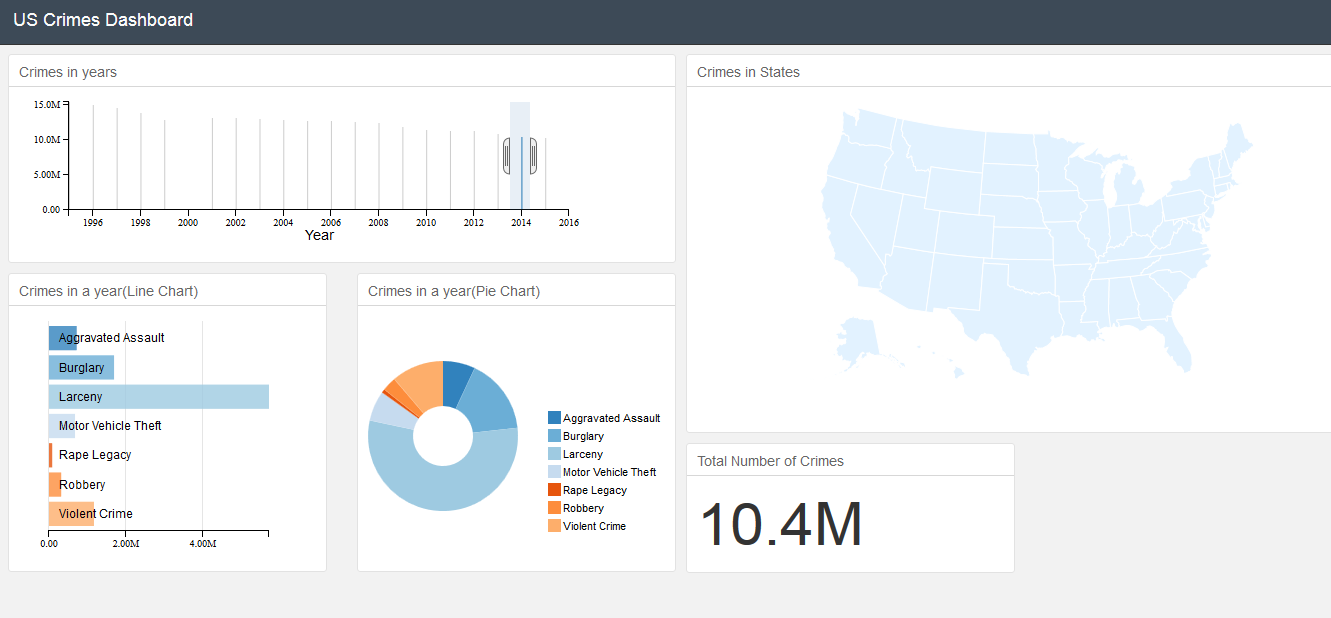
This dashboard gives an overall view of crimes and also under crime category for a period of 20 years

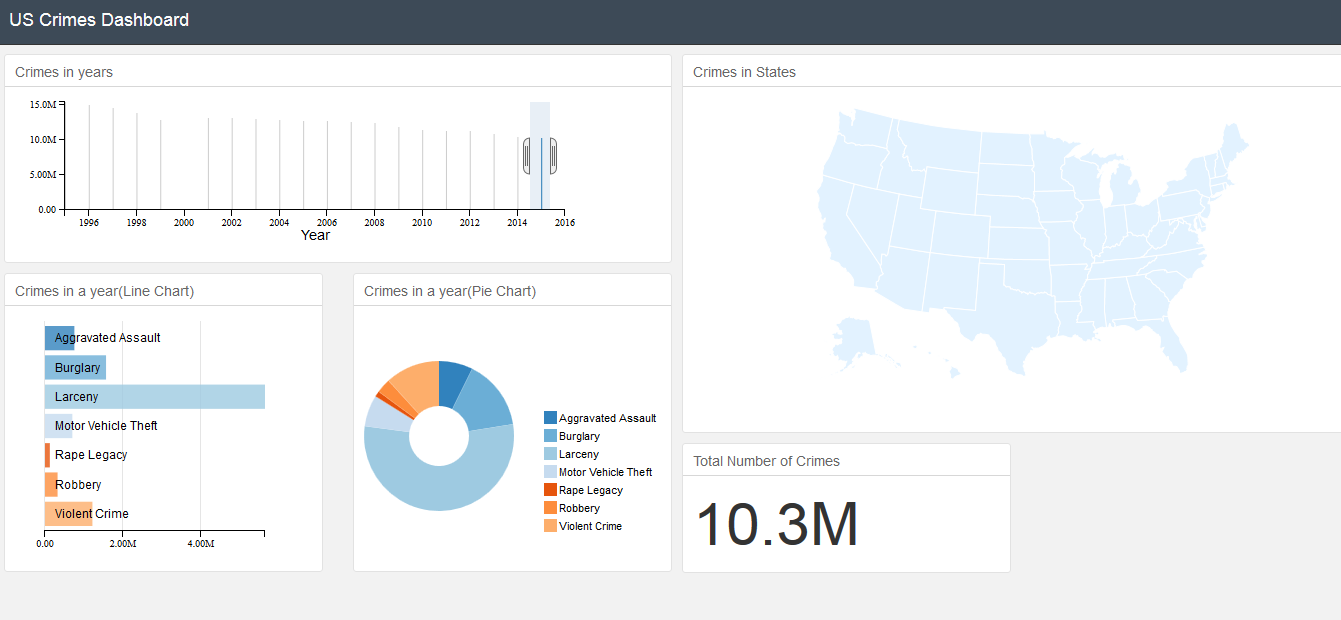


Total Crimes in 2015 is **1.36M in California**

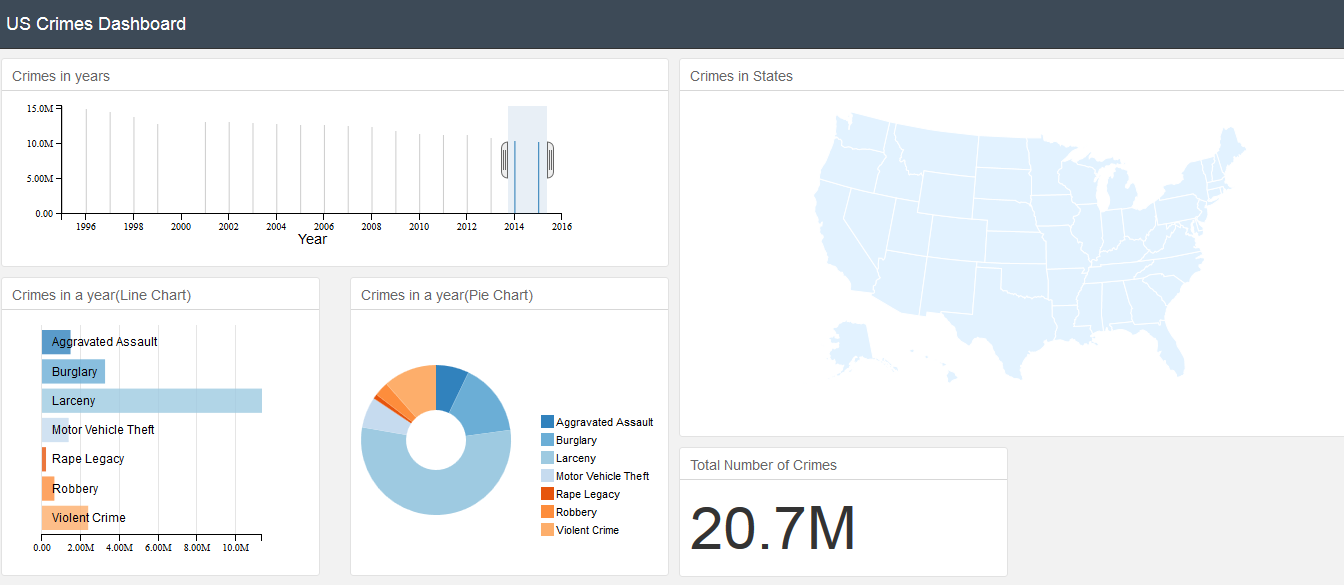


Total Crimes in 2015 is **1.25M in California**

Total Crimes in 2015 is **10.4M**



Total Crimes in 2015 is **10.3M**



Addition of Crimes in years (2015,2014) from all states in US is 20.7M

This interactive dashboard gives various insights on Crimes in US, from this dashboard you can see a trend on how crimes is having a **decreasing trend** from year 1996 to 2016

**Conclusion**

1. We have noticed that the variables have trends in the crimes. That is, more than 90% of **States** have the crimes that have the **Population < 15 Million** and **GDP < 800 Billion** and **Literacy Rate < 16%**.
2. ***Unemployment Rate*** is not a good indicator for the Crimes this concludes in our visualization that there no trends associated with the all other variables(attributes) in the dataset.
3. We have observed that all features are ***positively*** correlated and there exists a ***Strong*** correlation between **Population vs Total Crime** Scatterplot i.e., the state that as higher the population has the higher the crime rate. And also, **Total Crime vs GDP has a Strong correlation** i.e.,the state with the higher GDP has the higher crime rate**.**

**References:**

1. [**https://hub.arcgis.com/datasets/a2a8826c4ba44f89a684000fe7c05b8c\_0/data**](https://hub.arcgis.com/datasets/a2a8826c4ba44f89a684000fe7c05b8c_0/data)
2. [**https://www.worldatlas.com/articles/us-literacy-rate-by-state.html**](https://www.worldatlas.com/articles/us-literacy-rate-by-state.html)
3. [**https://www.bls.gov/opub/ted/2016/u-3-and-u-6-unemployment-by-state-2015.htm**](https://www.bls.gov/opub/ted/2016/u-3-and-u-6-unemployment-by-state-2015.htm)
4. [**https://www.santarosa-lawyer.com/united-states-crime-map-2017/**](https://www.santarosa-lawyer.com/united-states-crime-map-2017/)
5. [**https://crime-data-explorer.fr.cloud.gov/downloads-and-docs**](https://crime-data-explorer.fr.cloud.gov/downloads-and-docs)