# GUN VIOLENCE IN THE USA

**ABHISHEK CHANDRAN**
abhishek.chandran@utah.edu
**uID: u125771**

**VENKATESH SHARMA**
venkatesh.sharma@utah.edu
**uID : u1061421**

**PROJECT REPOSITORY:**
https://github.com/abhishekchandran/us_mass_shooting_analysis

---

## BACKGROUND AND MOTIVATION

Gun violence in the United States results in tens of thousands of deaths and injuries annually. The US has by far the highest number of privately owned guns in the world. It's one of the few countries in which the right to bear arms is constitutionally protected. The primary reason to choosing this project is motivated by mass shooting events that have been taking place regularly for instance an shooting event that took place on the night of October 1, 2017, where a gunman opened fire on a crowd of concertgoers at the Route 91 Harvest music festival on the Las Vegas Strip in Nevada, leaving 58 people dead and 546 injured.

## PROJECT OBJECTIVES

The main objective behind the project is to analyze how the proportion of mass shootings in the US has changed across past few decades. This is intended by developing an interactive visualization where the analyst will have the ability to get "detail on demand" stemming from overall view. This project will answer questions like which year wise number of deaths due to firearms? Safest and worst hit states in US and show the relationship between the number of deaths due to gun in a state and how strict are gun laws in that particular state i.e. we are trying to answer if gun laws and no of deaths are related to each other.

We look forward to this project as a learning opportunity where we can grow our skill sets academically and professionally by showcasing the visual design skills with the use of JavaScript and its D3.js library. We would learn about the implementation of best visual encoding techniques. This project will also serve as a platform to use the efficient Data-visualizations and can then be applied to diverse set of Data where one need to make comparison of any event over the years and spread across different geography. Also, this project experience will help us understand better about what type of visualization to choose to convey the insights out of any data set clearly and appropriately.

## RELATED WORK:

We decided to take up topic on gun violence in United States after so many of them were being reported frequently. We have been reading about number of people who been dying due to gun violence and controversial debate on gun laws in different states. With multiple data sets available in Kaggle, we decided to go ahead and do some exploratory analysis on data available from 1966 to 2016.

## DATA SOURCE

We are collecting mass shooting data from kaggle's datasets and state respective gun law details from statefirearmlaws.org.

[1] https://www.kaggle.com/carlosparadis/stanford-msa/data

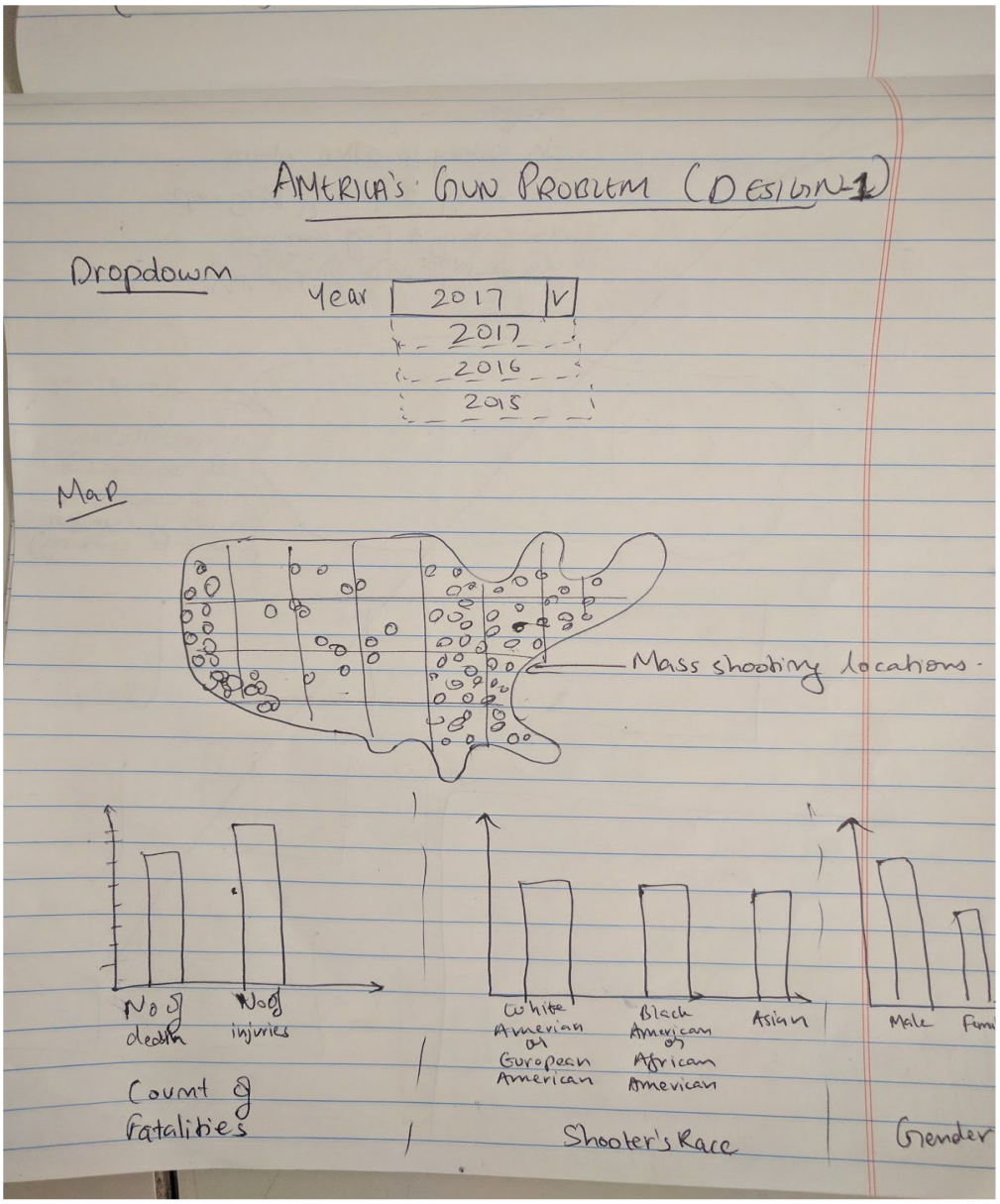
[2] https://www.statefirearmlaws.org/table.html

## DATA PROCESSING

The dataset we got from kaggle and state firearms website have all the attributes we would need to implement necessary visualisations.Though we had to segregate and nest it according to our chart, we don't see any particular need to do any kind of data processing.

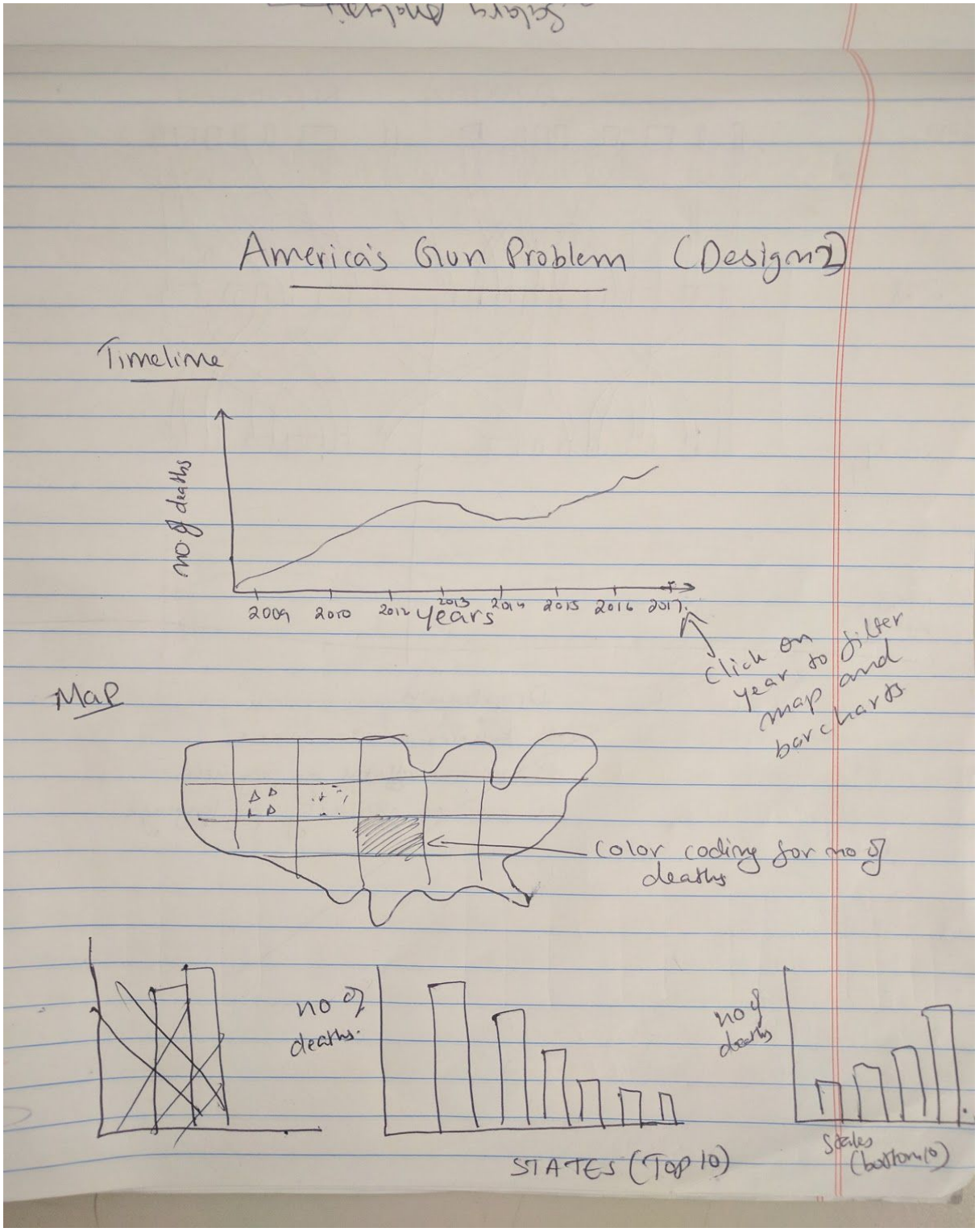
# DESIGN EVOLUTION

## DESIGN-1



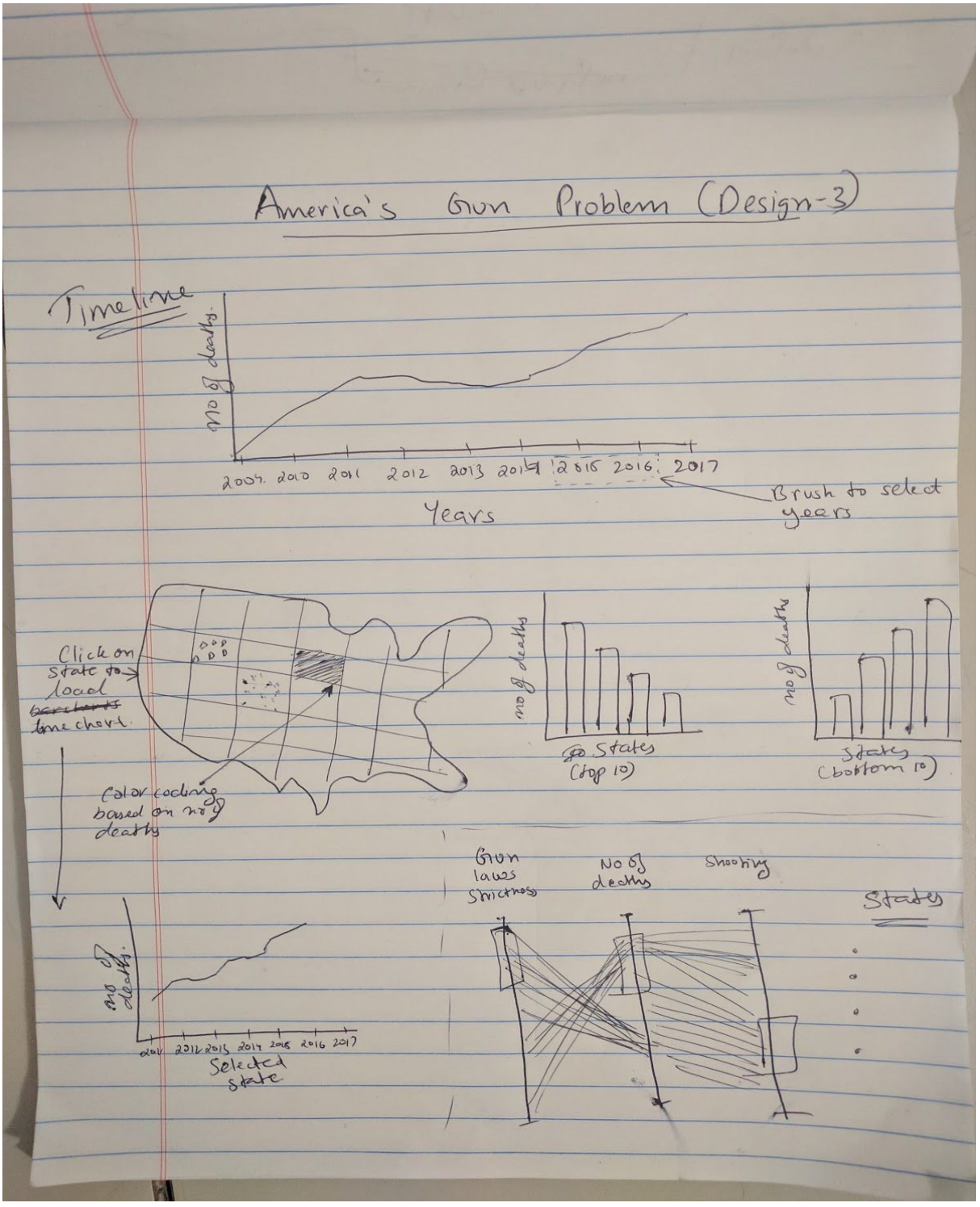

In this design we thought of providing the user a filter option through drop-down, so that he can select the year for which he wants to view different charts. It has a map which shows the density of shooting that have taken place in each location and is represented by circles. Based on year selected through drop-down, three bar charts will be rendered which will show count of fatalities, shooter's race and gender.

**DESIGN-2**



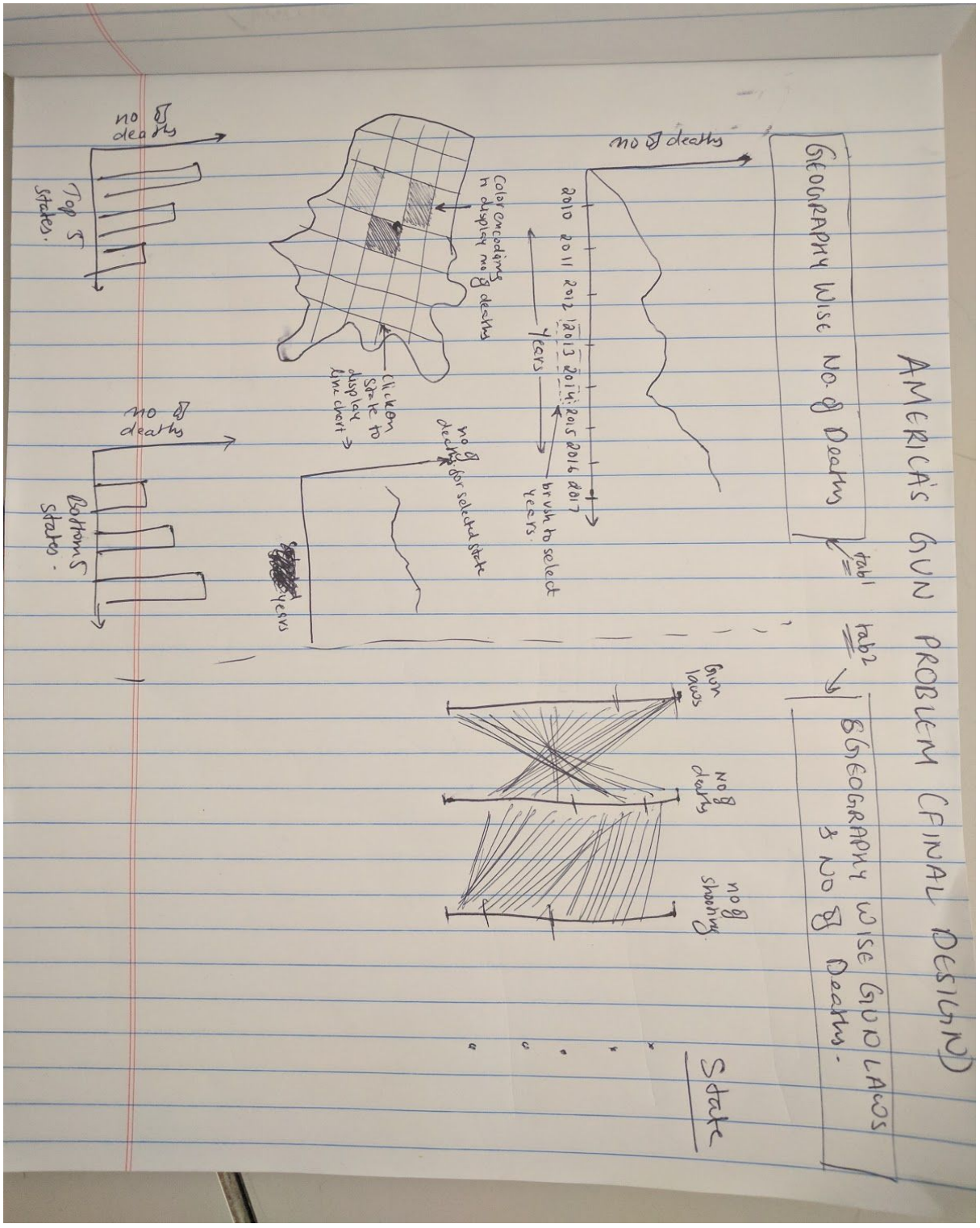

In our second design, we decided to get rid of the filter through drop-down and show a timeline story showing overall picture over the years, and filtering can be done by clicking on the labels on x-axis i.e. years. We changed the map from previous design to a map that shows color coded states based on number of deaths. We also replaced the attributes of the bar graphs with top ten states and bottom ten states w.r.t number of deaths.

**DESIGN-3**



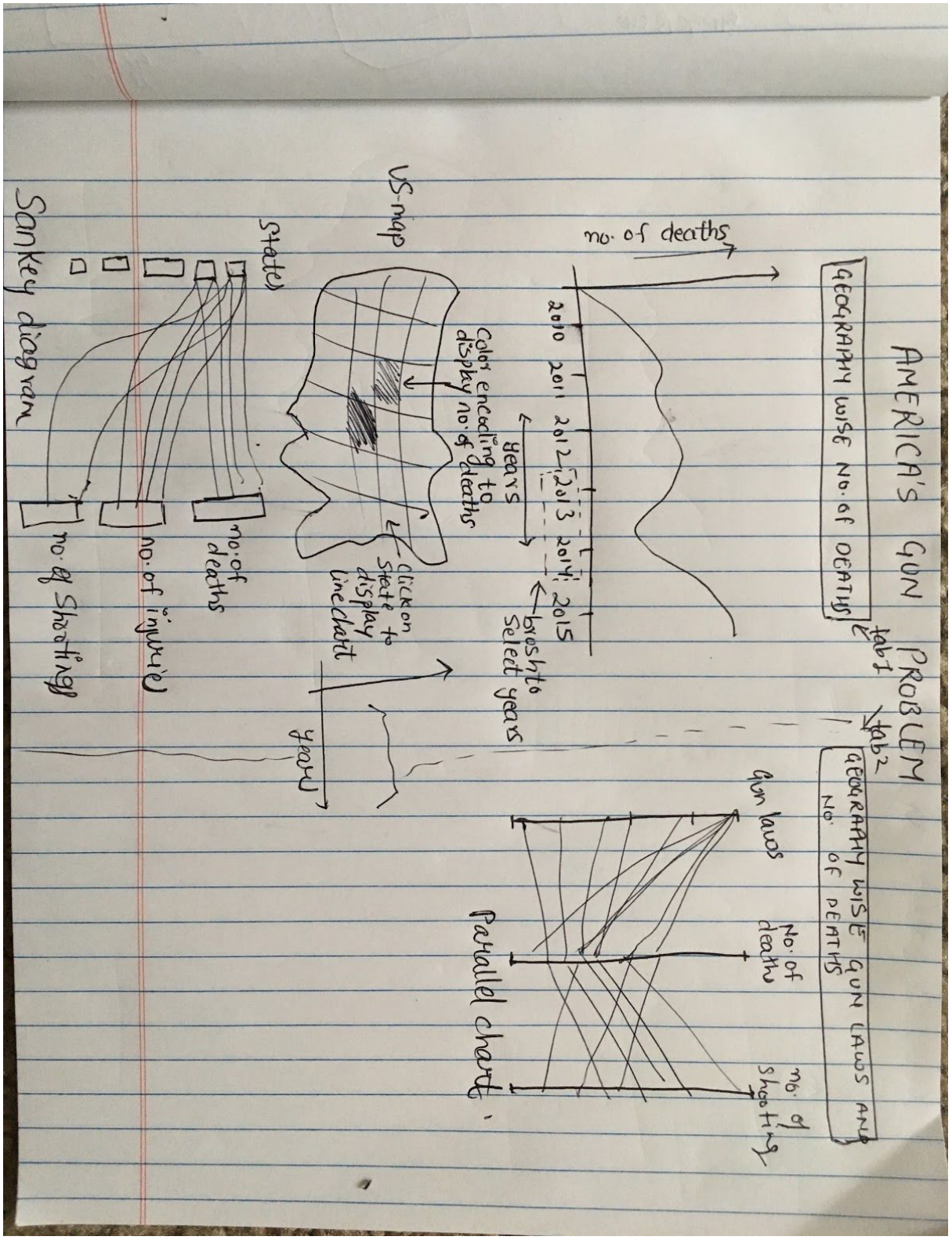

I our third design we have replaced clickable years on timeline with brush selection i.e. now user can select visualisation for multiple years. Clicking on states in map displays a line chart that shows trend in that state over the years. We have added one more chart, parallel chart, which has three columns each depicting gun law strictness, no of deaths and no of shooting, it can used to depict relationship between law strictness in each state and no of deaths.

**DESIGN-4**




Here we have segregated parallel chart and all other chart into two different tabs, as it gives a more clean view without much cluttering.This is the best taken from all the above designs.

**DESIGN-5**




This is final design we came up after peer feedback session. Here we have replaced bar charts with sankey graph, where we have state wise number of deaths, number of shootings and number of injured.

## MUST HAVE FEATURES:

1) Timeline showing no of deaths over the years, with brush selection for subsequent graphs.
2) Map with states color coded based on intensity of deaths.
3) Line chart showing death trend in selected state over the years.
4) Sankey Graph with number of deaths, shootings and number of injured.
5) Parallel chart showing relationships between state gun laws and no of deaths due to gun and no of shootings.

## OPTIONAL FEATURES:

1) On map show intensity of deaths with more precise latitudes and longitudes i.e w.r.t to cities
2) Show country stats in a table that can be sorted on multiple parameters
3) Account for changing gun laws over the years in each state.

.

## PROJECT SCHEDULE:

| Date | Group Member | |
|---|---|---|
| | **Abhishek** | **Venkatesh** |
| | | |
| 3rd Nov | Timline Graph | US Map |
| 9th Nov | Bar charts | Line chart |
| 17th Nov | Parallel Chart | Brush on all graphs |
| 23th Nov | Highlight citites on map | Show data in table |
| 29th Nov | Create Website | Complete Process Book |

## QUESTIONS WE ARE TRYING TO ANSWER:

1) What is the trend of deaths due to gun violence in US in past forty years?
2) In a particular year or set of years, which all states in US have been severely affected due to mass shootings w.r.t. number of deaths and injuries.
3) In a particular state, what is trend that has been taking place over the years?
4) What is the difference in magnitude of fatalities due to violence among different states in America? Which are the most affected and least affected states?
5) Do stringent gun laws affect the magnitude of violence?

During initial phase of the project we just trying to find trend over the years in each state, and as we progressed project evolved to a point where we are thought of finding relationship between gun laws and magnitude of violence.

## PEER FEEDBACK:

Peer feedback was conducted on 2nd November 2017, Vishal Pandey, Tanveen and Pranav reviewed our project and gave some valuable feedback and raised some valid questions which helped us in improving our visualisations.

Some of the questions that were discussed are:
Q: How are you going to segregate and valid data from multiple data sets?
A: Though all the attributes we need can be found from couple of csv files, we would have to create another file with all segregated data. We have written a python script for that activity.

Q; Is there any other innovative way where you can show magnitude of violence w.r.t. to each state?
A: Earlier we planned to show this data with bar charts, but after discussing this with peers and researching more about the different visualisation techniques we found sankey chart to be apt for representing state wise magnitude.

Q: Does visualization follow principles used in class?
A: We have put in our best effort to follow principles taught in class, like going those visualizations that are more effective in conveying all the information for e.g. visualizing color coded geographies on map, not going for pie charts(due to high number of attributes involved) and sankey charts for analysing the trend.

Q: How are going to make map more interactive?
A: We plan to highlight state on sankey graph when used clicks on a state in the map. We also intend to show summary stats through tooltip on hovering over state in the map.

Q: How will the user select year for which he wants to visualize the data?
A: We plan to implement a brush on timeline chart, where user can select a year/set of years, and map and sankey chart will be updated according to it.

## EXPLORATORY DATA ANALYSIS:

In order to visualize the data, we intend to use the following visualization techniques/charts:

**Timeline Graph:**
We have implemented a timeline graph which shows trend of no of deaths over the years i..e in last 40 years. User will be able to select the year for which he wants to visualize rest of the data through a brush on year's text display.

**Map:**
We will be using map to show regions which are most affected by gun violence, we intend to convey it through color coding i.e. more intense colors in areas which are most affected. User will be interact by clicking/hovering over the graph. Hovering will show overall stats and clicking will display line a chart which will convey trend over the years in a particular state.

**Sankey Graph:**
We intend to use sankey graph to show magnitude of deaths, shooting and number of injured w.r.t each state. It will show an overall picture of all states with more defined magnitude.

**Parallel Coordinates Graph:**
We have implemented a parallel coordinate graph to visualize relationship between state gun laws, no of shootings and no of deaths. We intend to learn if stringent state laws affects the violence due to gun. Brush has been implemented to select range of magnitudes.

## DATA PRE-PROCESSING:

We used the Number of Fatalities column from the data-set to measure the number of deaths in our timeline chart. We used R and Python to modify csv files which has only the number of fatalities associated with each year for the whole country. This was achieved by implementing the logic of summarizing the value of fatalities grouped by year for each state in our data-set. In our data folder for the application we have <state-name>.csv file which are created for generating the state specific timeline chart. This file contains “id”,”year”,”sum” column.

We have also created a cumulative_shooting_state.csv file which is the representation of number of fatalities per year for every state.

For making the parallel coordinate chart we also needed a modified data-set which represents the relation between the state name, the total number of gun-laws in those states, the number of shooting occurrences and the number of deaths/Fatalities associated with that state. The R implementation code can be found under the scripts folder of our code submission.

## IMPLEMENTATION

**Line Chart** :

The project starts with the implementation of a Line chart which serves as a timeline of how the number of fatalities/deaths changed over the years. On the x-axis of the chart years are represented in the increasing order. The X-axis is made out of the ordinal scale as the data set consisted of discrete years. On the Y-axis we used the linear scale to represent the number of death/fatalities over the year. The intention of making this chart is to show the trend line over the course of years and to answer the questions like whether the number of deaths has increased/decreased over the years. This chart also has a feature of brush selection associated with it. We can select the years to trigger the color change in the US-map. This brush makes an array of years which are selected by the user and fetches the data associated with those years to aggregate them and represent as a color coded US-map.

**Map**:

Just below the timeline chart we have placed the US map to display the effect of interaction with the timeline chart by the means of brush element. This chart has two objective which are as follows. The first objective is to display the color coded intensity of deaths/fatalities over the states for all the years from 1966-2016 aggregated together, this feature is represented as soon as the page loads. The second one is the to display the color coded intensity of the deaths/fatalities for the years selected by the brush element. The intention behind this chart was to see how the intensity has changed over the years for all the state. The changes in the chart triggered by changing the brush value gives us an idea of the states where this change has been drastic and where this change has been minimal. On hovering over the states on the map we can see the name of the state and therefore can make a decision to select that state by clicking on it. Clicking on any state will trigger another line chart which shows trend of the fatalities with their year of occurrence.

**Sankey Diagram:**:

This visualization shows the relation between the number of deaths, number of shooting and number of gun laws for the selection of states. This diagram is also triggered by clicking on the US-map for the clicked state. Also, user can select multiple states by checking them from “Add More States to Sankey” button. A search functionality is also included for checking the states for visualization. The intention behind this visualization is to answer questions like what is the effect of number of gun laws in a state on the number of shootings and the number deaths/fatalities associated with it. Whether the number of gun laws in any state affects the frequency of shootings and its associated fatalities/death. This can be clearly observed from

the diagram as the state with more number of gun laws tend to show less number of shootings and therefore lesser the number of deaths.

**Parallel Coordinates:**

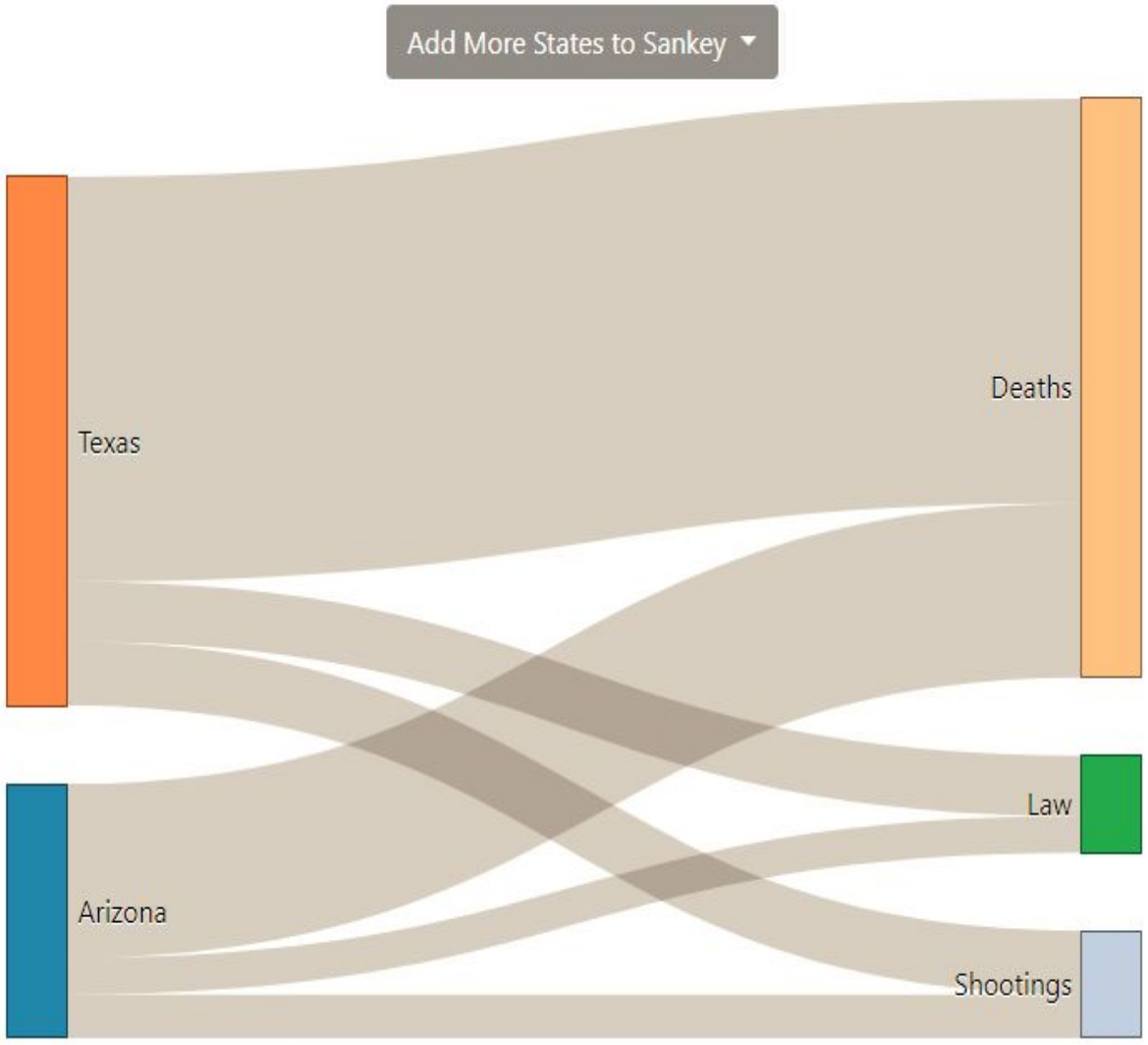
It is used for visualizing high-dimensional geometry and analyzing multivariate data. User can select the range of number of total gun laws and can observe the associated links with the number of shooting incidents and the number of deaths associated with it. This is therefore a powerful visualization to answer that more the number of gun laws the lesser are the shootings and as a result of which lesser is the number of deaths. This chart is made more interactive by linking a summary table to it. This summary table is updated as soon as the user changes the brush selection or moves the selected brush. The table gives us the name of state which falls under the range of selected gun laws and gives the data like number of gun laws, number of shootings and number of deaths associated with that state. The intent behind this diagram is to give the quantitative comparison between the number of gun laws and its associated incidents.

## DESIGN EVOLUTION:

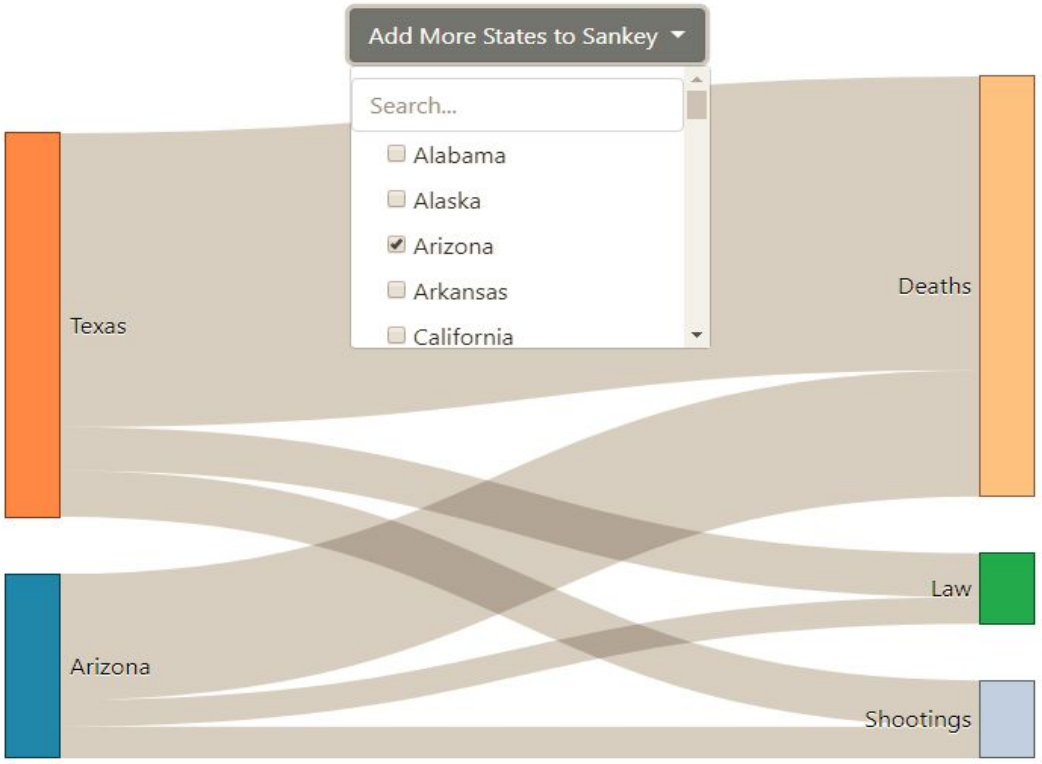
Initially we planned to show line chart which shows the timeline of events, a map with magnitude of death color coded, a sankey graph with all states related to no of deaths, no of shootings and law strengths and a parallel coordinates chart with with relationship between magnitude of events and state law strength. But during implementation we learnt that we need to we should do more to provide interactions for the user. So we introduced bunch of functionalities like filter for sankey diagram, brush for selecting multiple years and a summary table for viewing data selected from parallel coordinates.

**Filters:**

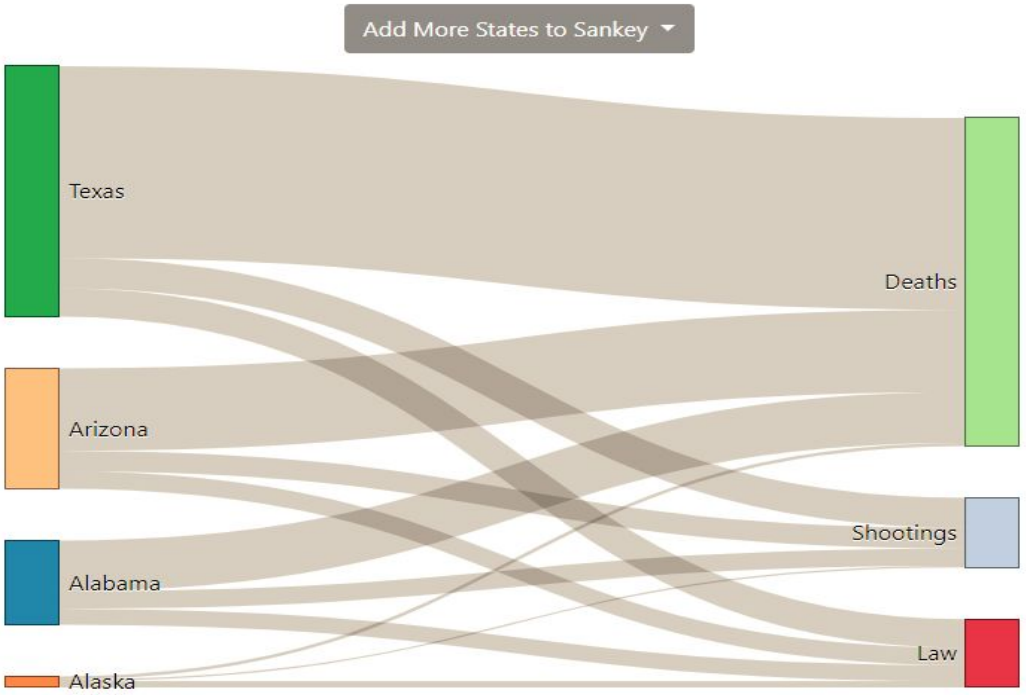
Earlier design did not have any filter on sankey diagram, but after implementing it we learnt that diagram was too big for displaying all fifty states, and it was not proving to be a pleasant experience for the user. So, in order to improve user interaction and present visualization in more aesthetic way, we implemented dropdown with all the states listed.




If sankey has two states included then, clicking on “Add more states to Sankey” will give a dropdown that can be used to select multiple states.

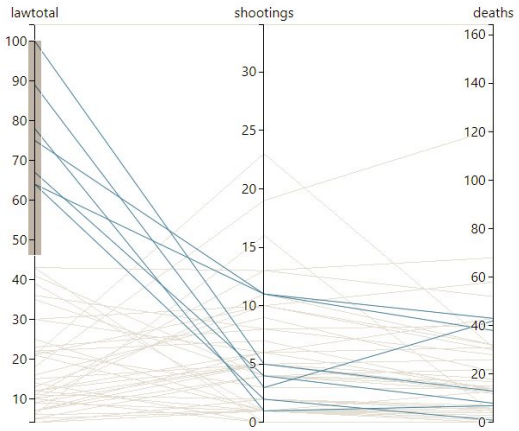



Clicking on Alabama and Alaska along with Arizona and Texas, a new sankey diagram will be displayed with Alabama, Alaska, Arizona and Texas.



## Parallel Diagram:

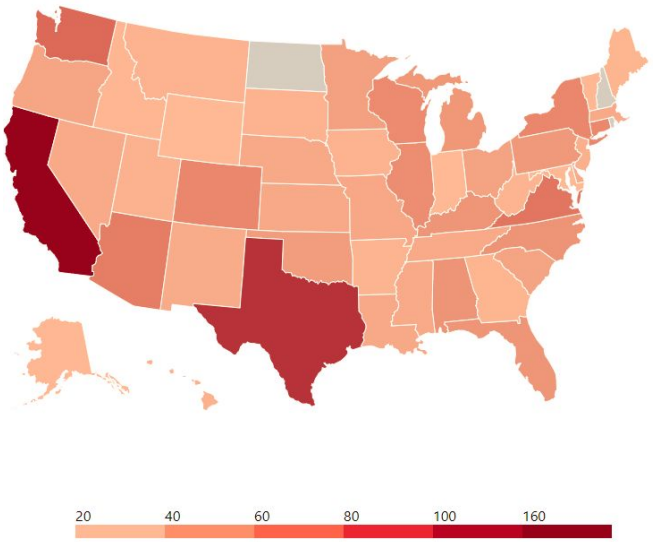
Initially we intended to just show the parallel coordinates, later on in order to make it more informative, we added a table with all the summary data. In parallel coordinates user can use brush over all three coordinates and corresponding filtered i.e. state selected through link will be displayed in the table. Parallel coordinates graph is used to analyse how an attribute behaves w.r.t another attribute. It can be used to analyse multiple attributes as well.

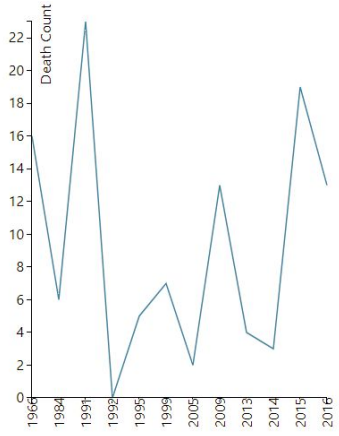



| State | Law | # of Shootings | # of Deaths |
|---|---|---|---|
| Connecticut | 89 | 3 | 42 |
| Hawaii | 78 | 1 | 7 |
| Illinois | 64 | 11 | 38 |
| Maryland | 64 | 2 | 1 |
| Massachusetts | 100 | 5 | 13 |
| New Jersey | 67 | 4 | 8 |
| New York | 75 | 11 | 43 |

## Map:

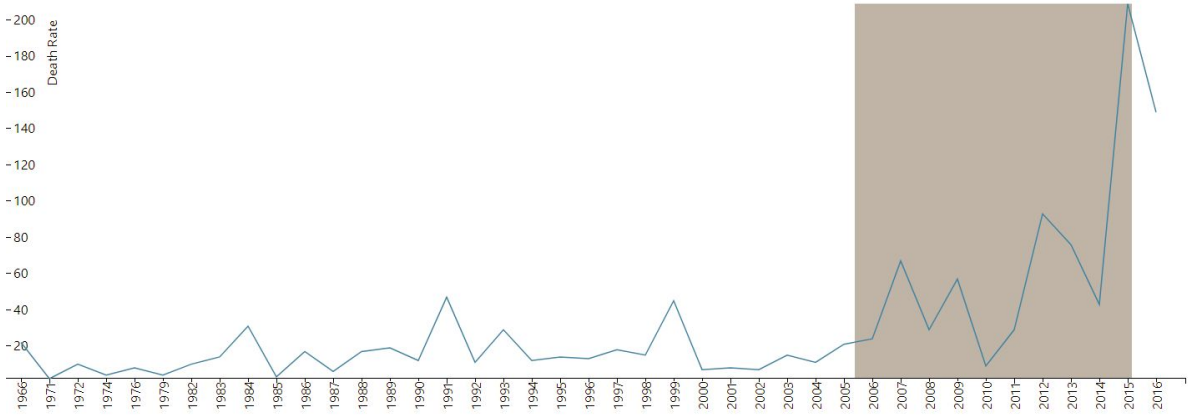
Map is used to view how each state have fared as compared to others. Color intensity keeps on increasing as no of deaths increase. Variation in map color can be seen by moving brush over the time line graph as shown in next section. Clicking on each state draws a line chart showing trend over the years in that state.



**Timeline Graph:**

Timeline graph shows how events have unfolded over forty years, how the overall trend has been w.r.t to every recorded year. User also have an option to use brush over the timeline graph and in turn it creates filter in the map. In this graph we have only considered number of deaths and not no of injuries or shooting, since this has already been covered in previous charts.




## EVALUATION:

We have tried our best to answer the questions we had when we started implementing this project, but we believe we would have had more substantial evidences if we had larger data set. Some of the insights we gained from this project are:

1) No of deaths have been increasing over the years due to mass shootings and it has risen substantially in past couple of years.
2) US citizens holds more than 50% of guns w.r.t whole world and number of deaths, injuries and shootings are more due this.
3) Some of states have stricter gun laws and in those states, number of fatalities have been less as compared to other states.
4) States with lenient gun laws have suffered the most, as number of dead are more in these states.
5) 60% of suicides are committed using a gun.

Some of the shortcomings in this project are:

1) Dataset is not that exhaustive to provide a substantial evidence.
2) We could have used suicide and individual crimes caused by gun in order to analyse the intensity.
3) Gun laws measurement can be made more definite by considering how they are calculated and what all things are taken into account.
4) Situation in US can be compared to different countries and how stringent are gun laws and number of deaths related to it

## REFERENCES:


[1] Carlos Paradis. Stanford Mass Shootings in America. Kaggle

[2] State Firearms Laws : Website

[3] German Lopez. America's Unique Gun Violence. Vox Article

[4] Kara Fox. America's Gun Culture. CNN Article