Visualising K-12 School Shooting Database

Visualisation Coursework

**Walidul Alam Ehab**

2257956

Coursework submitted for CS5803 Data Visualisation

Department of Computer Science Brunel University London

United Kingdom 24/04/2023

# Introduction

School shootings have become an all-too-frequent occurrence in the United States, with devastating consequences for the victims, their families, and the wider community. The rise of these incidents has prompted urgent calls for action to prevent and mitigate future shootings, and to develop strategies that can help to keep students and staff safe. To achieve these goals, it is essential to analyze data related to school shootings and to develop effective visualizations that can help to identify patterns and correlations in the data. Through this coursework, we aim to contribute to the ongoing efforts to improve the safety of our schools and communities, and to promote a deeper understanding of this complex and pressing issue.

## Data

This coursework focuses on the ”K-12 shooting” database, which provides a rich source of information on school shootings in the US from 1970 to 2022. The database is acquired from [the K-12 School shooting database.](https://www.chds.us/ssdb/) The database contains four data sets: INCIDENT.csv, SHOOTER.csv, WEAPON.csv, VIC- TIM.csv, interconnected by a unique incident identifier. For the purpose of this project, I will be focusing mainly on the INCIDENT.csv file, as it contains the most information about the incident itself, while also using a few variables from the weapon and victim data set. There are a total of 41 rows and 3487 rows after joining the 3 data sets. Description and domain about the variables which will be used in the project along with the ones from incident.csv are given below in the form of a data dictionary. Variables used in the project are in bold.

Table 1: Data dictionary

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Domain** |
| **Incident ID** | **incident unique identifier** | **string** |
| **Date** | **date of incident** | **Date** |
| **State** | **state of incident** | **location** |
| **Situation** | **Reason for shooting** | **nominal** |
| **Weapontype**(WEAPON.csv) | **type of weapon used** | **nominal** |
| **Injury**(VICTIM.csv) | **type of injuries involved** | **nominal** |
| Sources | source of information | string |
| Number news | Number of news reporting | integer |
| Media Attention | Scope of media attention | nominal |
| Reliability | reliability of source | ordinal |
| Quarter | school quarter | ordinal |
| School | name of school | string |
| City | city of incident | location |
| School Level | Level of school | nominal |
| Location | place of shooting in school | nominal |
| Location Type | General location type | nominal |
| During School | did shooting take place during school? | boolean |
| Time Period | school period of incident | nominal |
| Continued on next page | | |

**Table 1 – continued from previous page**

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Domain** |
| First Shot | Time of first shot | time |
| Summary | Summarised description of the incident | string |
| Narrative | Detailed description of the incident | string |
| Targets | Type of targetting | nominal |
| Accomplice | Presence of accomplice | nominal |
| Hostages | Hostage situation | boolean |
| Barricade | barricade applied | boolean |
| Officer Involved | Was an officer involved | boolean |
| Bullied | incident due to bullying | boolean |
| Domestic Violence | incident due to domestic violence | boolean |
| Gang Related | incident gang related | boolean |
| Preplanned | Shooting preplanned | boolean |
| Shots Fired | Number of shots fired | integer |
| Active Shooter | Was it active shooting | boolean |

## Persona and Questions

The user is a data analyst who is interested in understanding and analyzing patterns and trends of school shootings in the United States to gain a better understanding of the frequency and characteristics of these tragic events. There are two specific research questons:

1. Are there any commonalities or differences between school shootings that take place in different parts of the United States?
2. How has the frequency and pattern of school shootings in the US changed over the years?

## User Requirements

R1: To answer Q1, user needs to look at how the frequency of school shootings vary with different parts of the US.

* One suitable view would be to create a map plot with Latitude and Lon- gitude of cities in the x and y axes, while color-coding markers with the frequency of attacks.
* A year slider filter can be used to limit the range of years in view

R2: Question 1 may also be answered by comparing the prevalence of different types of weapons across different parts of the country.

* User can create a bar chart that shows the frequency of school shootings by state, color-coded by the type of firearm used.

R3: To answer Q2, user can look at how the frequency of shooting varied over time for different types of schools.

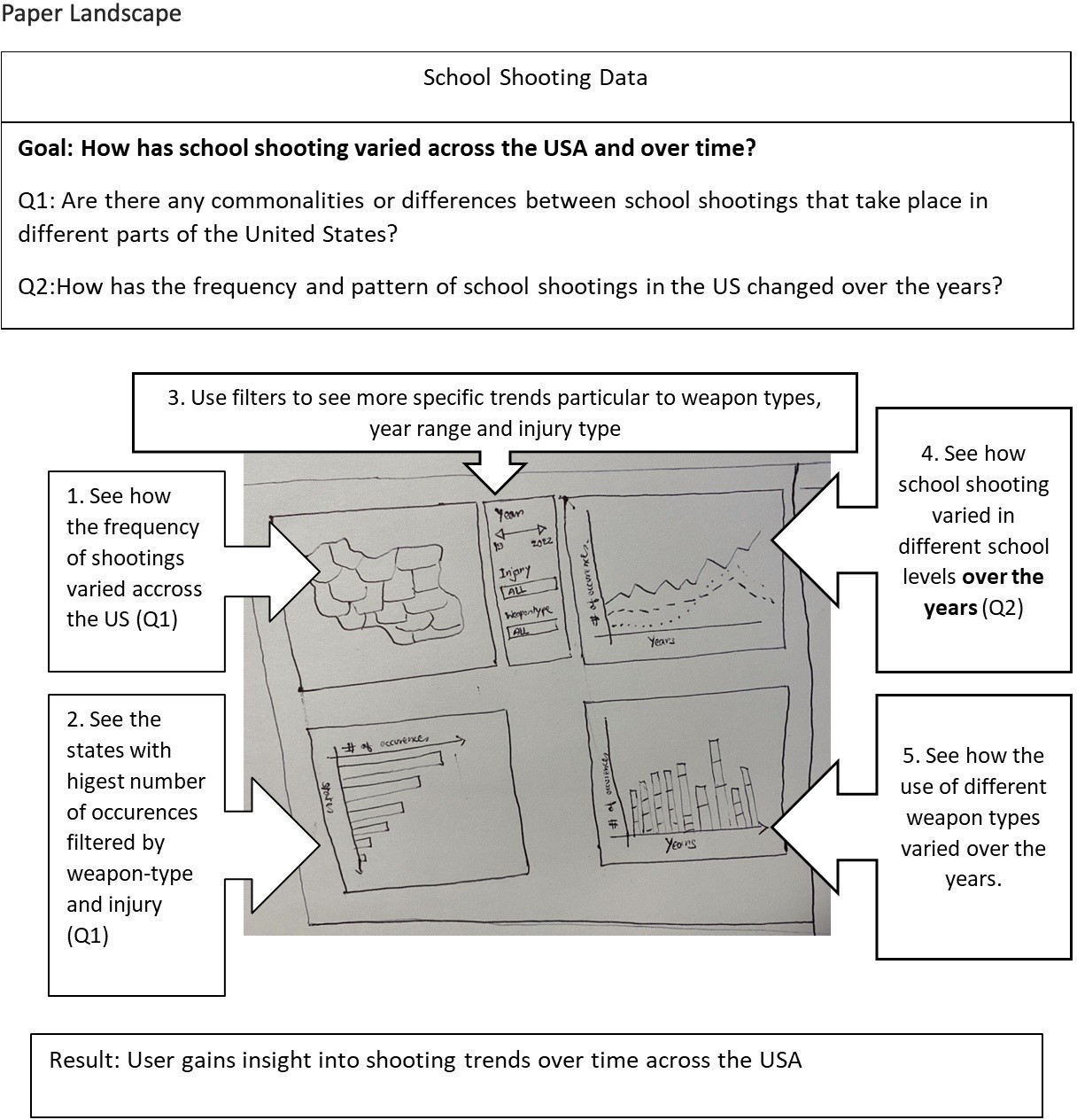
* User can plot a line chart with date in years in the x-axis while Frequency of attacks is in the y-axis.
* Additionally, a filter on the different types of schools can show how the frequency on each type of school changed over time.

R4: To answer Q2, user can also look at how cause of shooting changed, if any, over time.

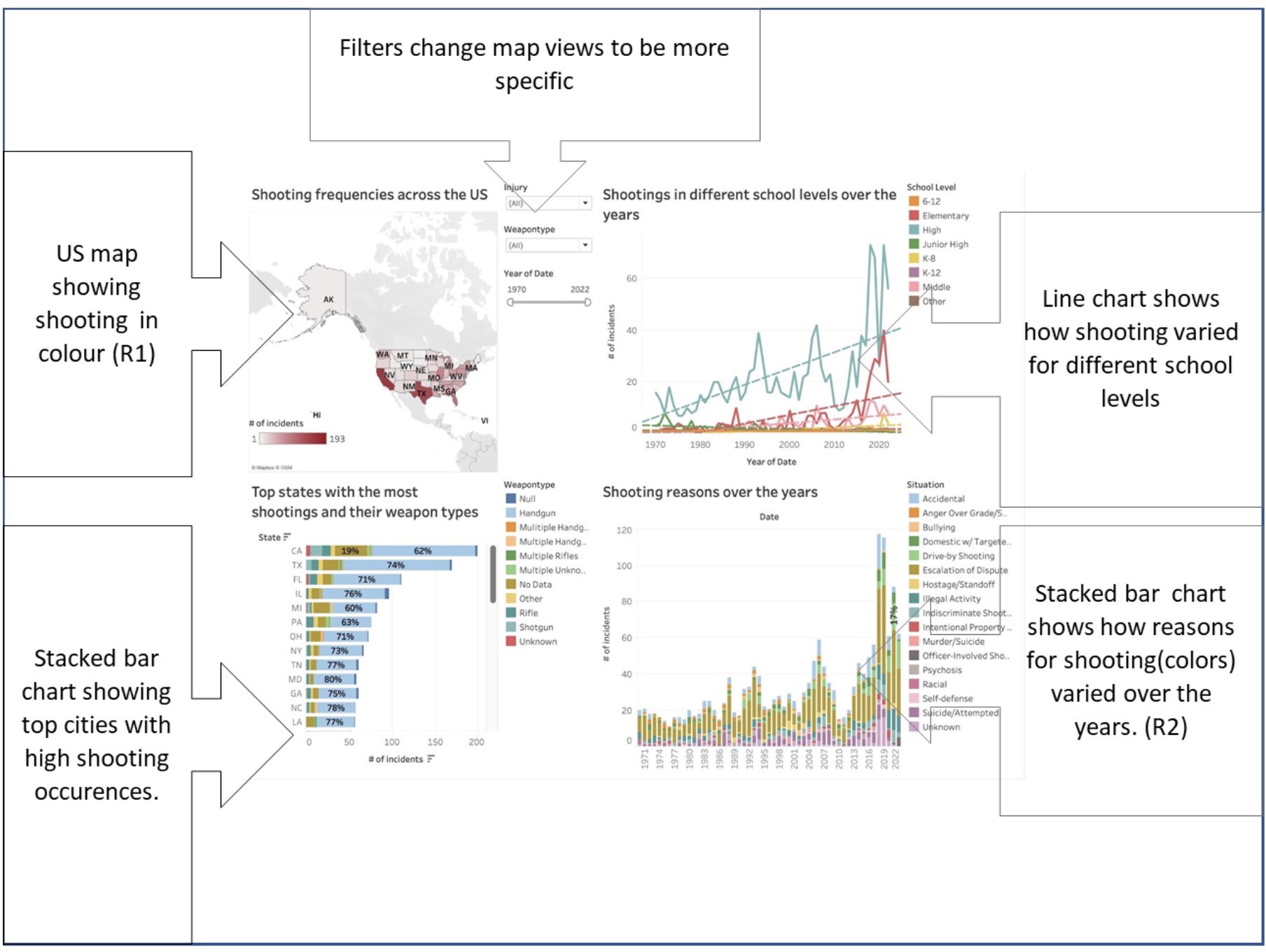
* User can plot a stacked bar chart with date in years in the x-axis while Frequency of attacks is in the y-axis. The stacks will show distribution of reasons for shooting involved every year.

# Design

* 1. **Paper Landscape**



* 1. **Design**



* 1. **Design choices justification**

Design choices made throughout the project was intended to be kept consistent with various image theories and encoding methods introduced in the CS5803 lectures. The visualisations were focused on looking at number of shooting incidents, which is a continuous datatype. The measure was visualised using color of varying saturation, size and lines (Lecture 2, CS5803 Lecture materi- als). Discrete nominal variables explored, such as weapon type, different school levels and shooting reasons, were encoded using different color hues, as this was an ideal practice in differentiating nominal values. Nominal comparisons, such as exploring incident distributions in different states was achieved through bar charts as this allowed distinctive comparisons(states) while maintaining relative quantitative comparisons(shooting incidents) (Lecture 3, CS5803 Lecture Ma- terials). Time-series data, i.e, shooting variations over the years, was achieved through line chart which helped visualise effortlessly how shooting occurrences traversed through time. Axes lines and labels were less focused compared to data lines to minimise non-data pixels and chart junk (Lecture 4, CS5803 Lec- ture Materials). When designing the dashboard, attention was given to align similar charts together, more specifically charts contributing to similar research questions were aligned vertically, and legends were placed side-by-side o help user readability.

## Difference(s) in initial prototype and final design

There was a change in the stacked bar chart which was used to fulfill R2 and answer question 2. Initially, I planned on creating a stacked bar chart where

colored stacks will be used to display how the weapon types varied every year. This seemed redundant to me when working towards implementation, as I found there wasn’t many noteable information to talk about and that I have already explored weapon types when answering Q1. I decided on using the colored stacks to represent reasons for shooting instead, as I believed this gave me and the readers a new perspective at exploring the data. I believe the change is effective at answering Q2 as there are more trends revealed.

# Implementation

## Tableau Implementation

### Preparing data source

The first part of the implementation involved preparing the datasets. As men- tioned before, the K12 SSDB includes three four data sets(INCIDENT, SHOOTER, VICTIM, WEAPON). My analysis mainly focus on the INCIDENT.csv dataset,

as it describes the most information about the incidents. However, features from other tables were also required, namely injury and weapontype from VICTIM and WEAPON tables respectively. This required joining tables using the com- mon incidentID variable as the identifier.

### ”# of occurences” variable

Joining of tables created a a lot of new rows with the same incident id. This is fixed by creating this new variable, by specifying a distinct count formula on the incident ID variable from the INCIDENT table.

## Chart I - Geo-map of US

The first chart is initialised by creating a new worksheet in Tableau under the same project. The state variable is used to assign latitude and longitude variables to the axes. This generates a world map. Next, the # of occurences variable is dragged to the colors marker in the Marks card. This showed the frequency of incidents as color on each state. The color palette is adjusted to resemble that of danger. Next, Date(YEAR), Weapontype and Injury(VICTIM) variables are dragged to the filter card one by one to add filters to the view. Weapontype and Injury are kept as single valued dropdown filters, while year is a slider which can select a range of years. The YEAR variable is assumed continuous to enable this flexibility.

## Chart II - Bar chart showing number of ocurences in different states

This is done by implementing a bar chart by assigning State to column and # of occurences in rows. The resulting bar chart is sorted in descending order using the x-axis. Next, Weapontype variable is assigned to color card to enable stacks and occurences variable is dragged to label card to show number of incidents in the stacks. This is then converted to percentage of total cells to show

the weapon type percentage in each state. Weapontype and Injury(VICTIM) filter is added to allow global filtering in dashboard.

## Chart III - Changes in number of incidents over the years for different school levels

This is implemented using a line chart by assigning # of occurences to Columns and YEAR measure as Row. This creates a line chart showing overall variations in the incident frequency over the years. Next, School Level variable is dragged to Color card to allow multiple lines representing different school levels. Trend lines are generated using Tableau’s analysis menu to show trends in for each line.

## Chart IV - Causes of shooting every year

The implementation is a stacked bar chart with Year in the row and # of occurences in the column. Reason variable is dragged to color card to enable stacks. Similar precedure as Chart II is done to display percentage of reasons in each bar.

## Dashboard

The dashboard is created using the four charts on a Letter Landscape size. Year, Weapontype and Injury filter is shared between the first two charts, while legends for each graph is placed alongside to ease readability.

# Walk-through

The walkthrough section will guide the reader to learn about the solutions or discoveries found in response to the research questions asked and also display the steps which brought about said discoveries. For revision, the research ques- tions are re-stated below:

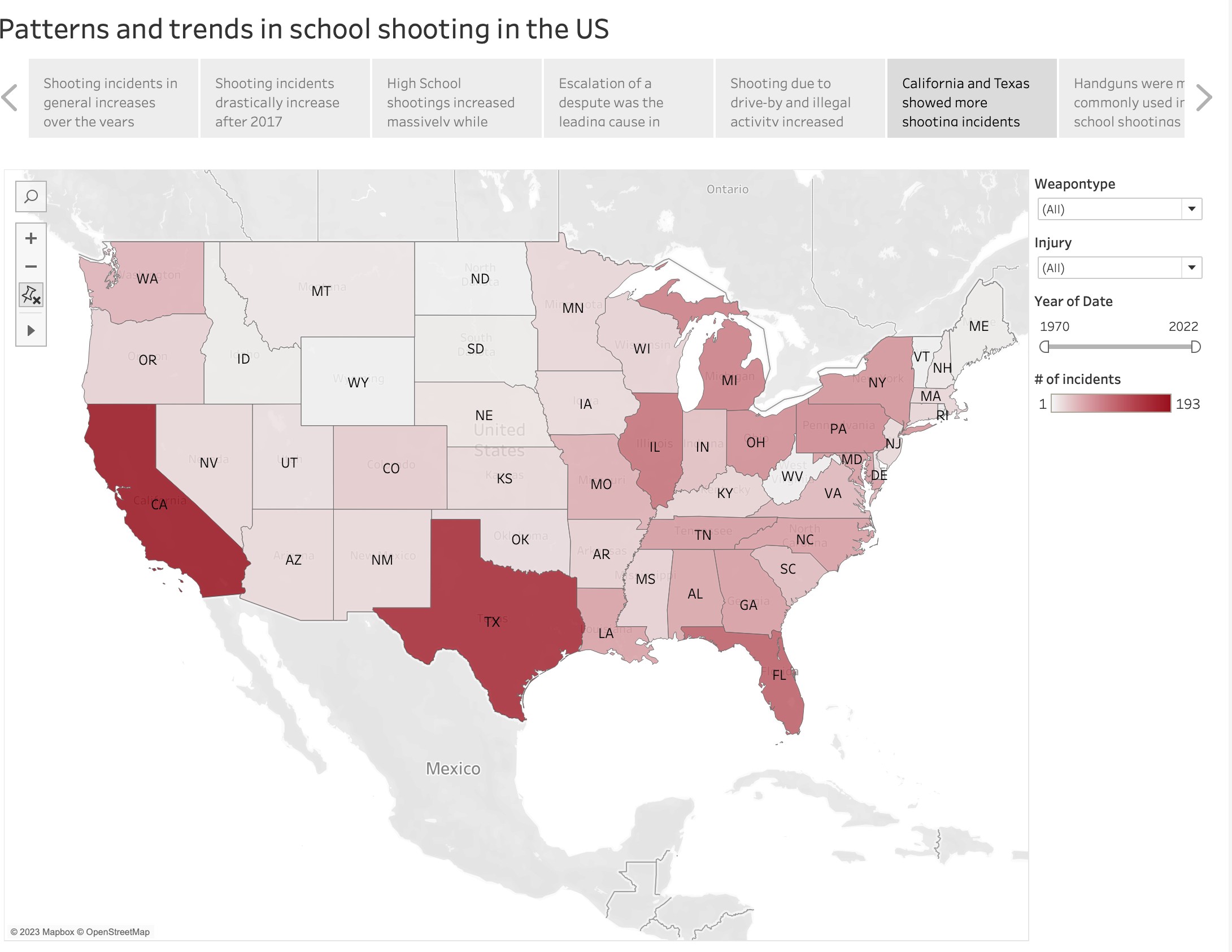
**Q1**:Are there any commonalities or differences between school shootings that take place in different parts of the United States?

**Q2**:How has the frequency and pattern of school shootings in the US changed over the years?

## Differences and similarities in school shootings across the US

**Discovery 1**: California and Texas showed more shooting incidents than any other states.

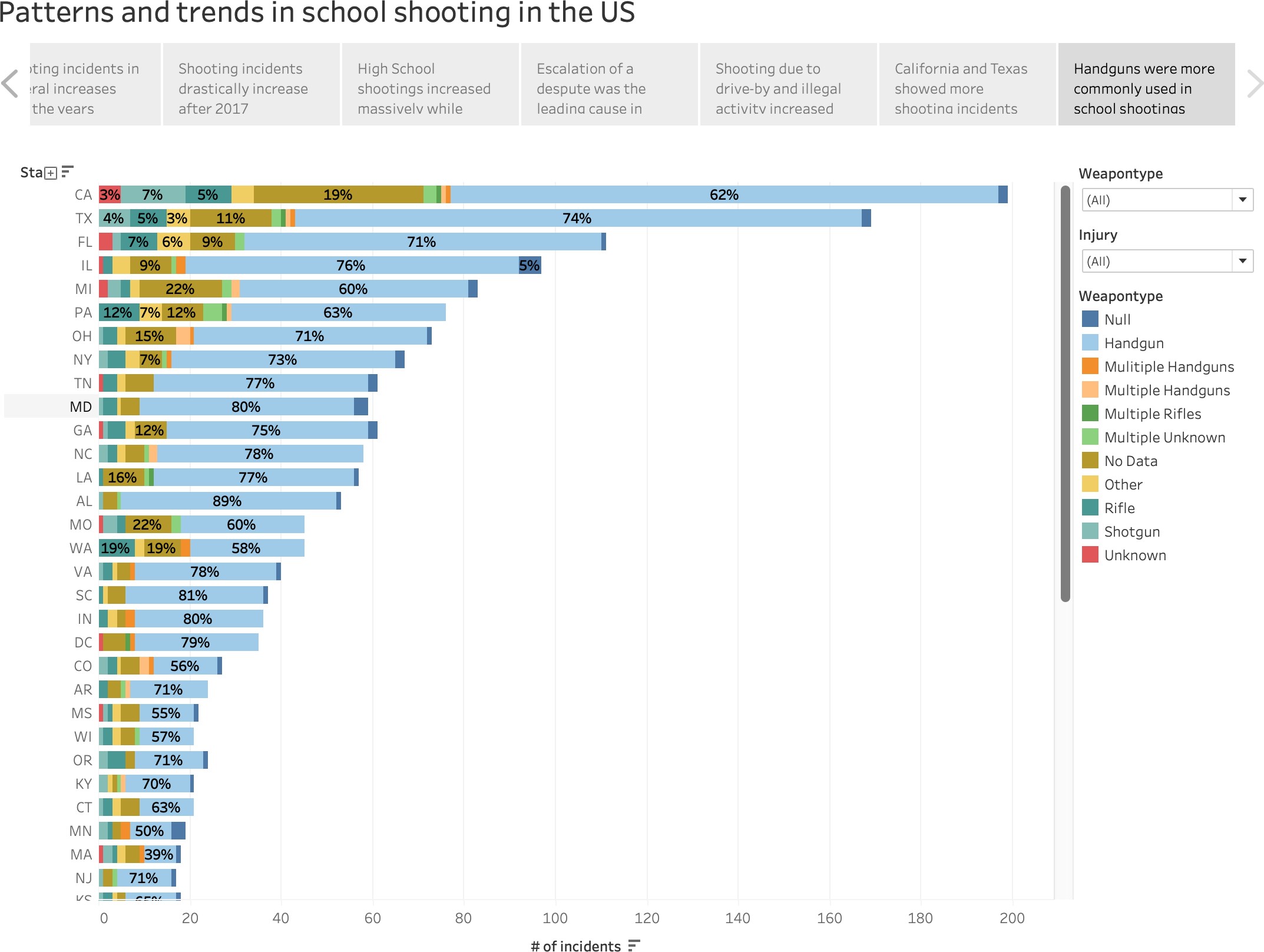
This was found from the mapping plot where states were colored according to the number of incidents that took place over the years. A same color scale(red) was used to highlight the frequency of attacks due to its continu- ous nature. The map revealed that California(CA) and Texas(TX) showed the highest number of incidents over the span of all years in record, which was evi- dent from the dark red shades. Color mark on other states were much lighter.



This was further confirmed from the second chart plotted to answer the question.

**Discovery 2**: Handguns were more commonly used in school shootings across all states.

Handguns were found to be the most common type of weapon used during shooting in all states. The horizontal bar chart which displayed the distribution of weapon type showed handguns to have the longest stacks in all the leading states in terms of shooting frequency.

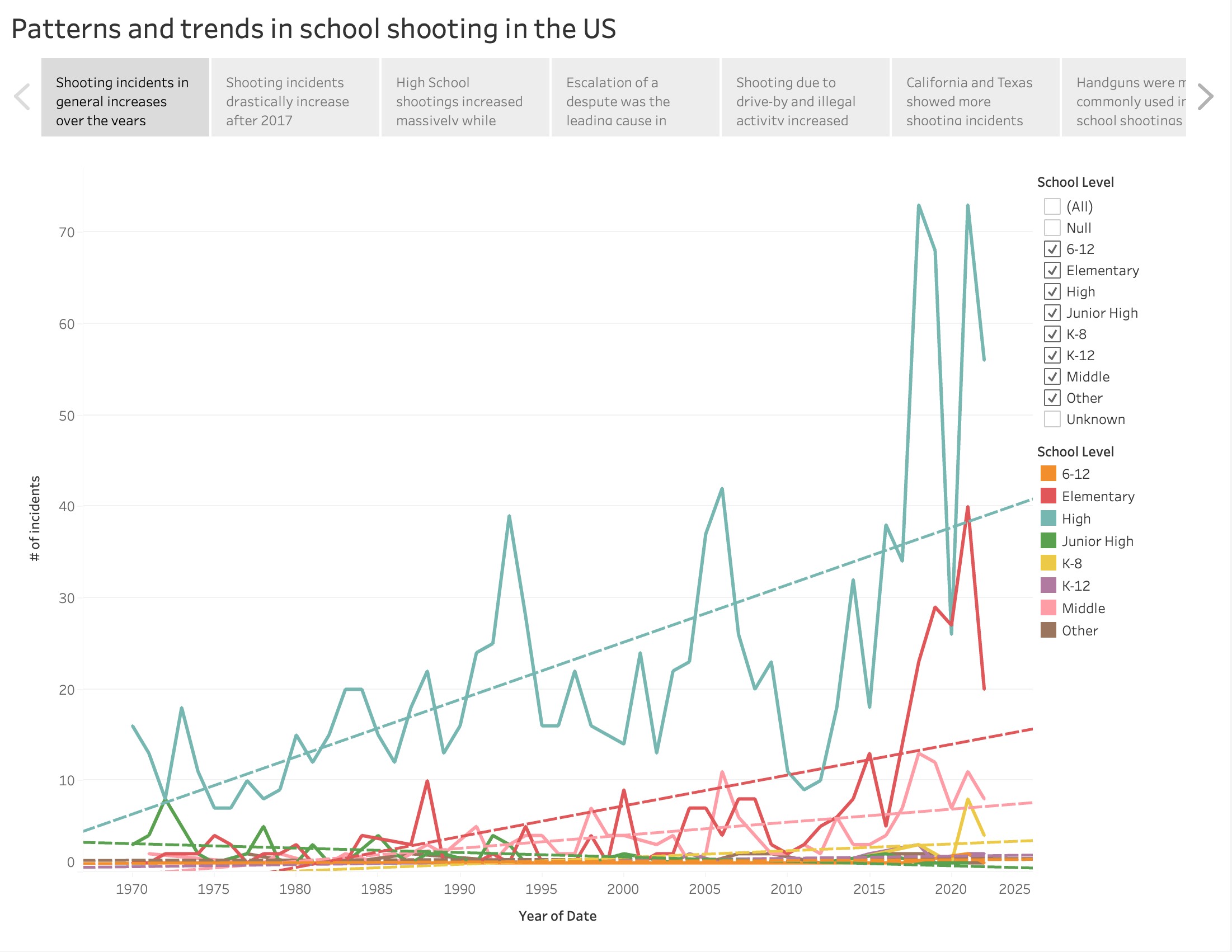


The chart also confirmed California and Texas being the most frequent locations of school shootings.

## Patterns in school shooting over the years

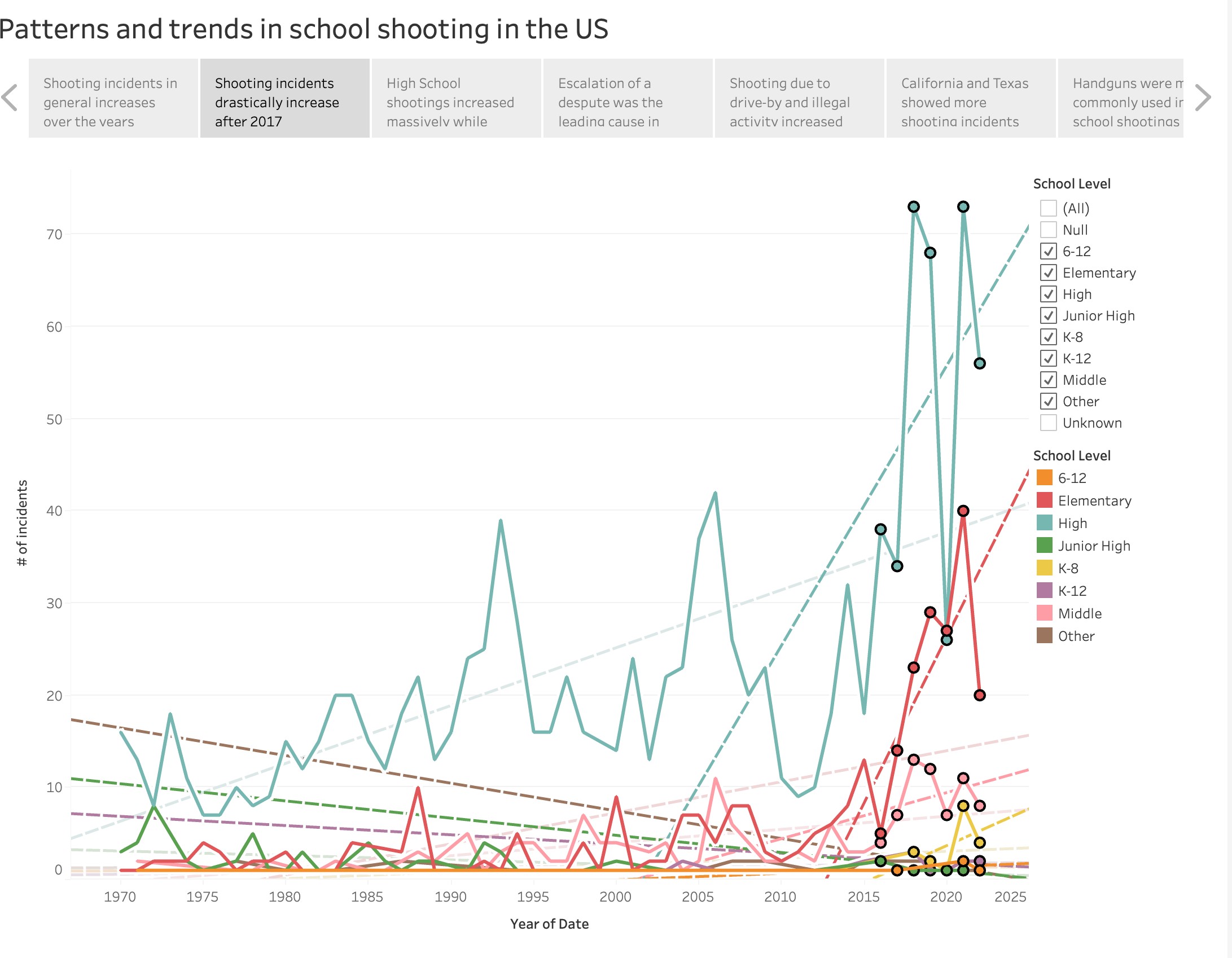
**Discovery 3**: Shooting incidents in general has increased over the years.

First impressions on the line chart of occurrences vs years revealed that shoot- ing frequencies have generally risen over the years. Most of the trend lines for different school levels has a positive slope, meaning shooting frequency has had an overall increasing trend.

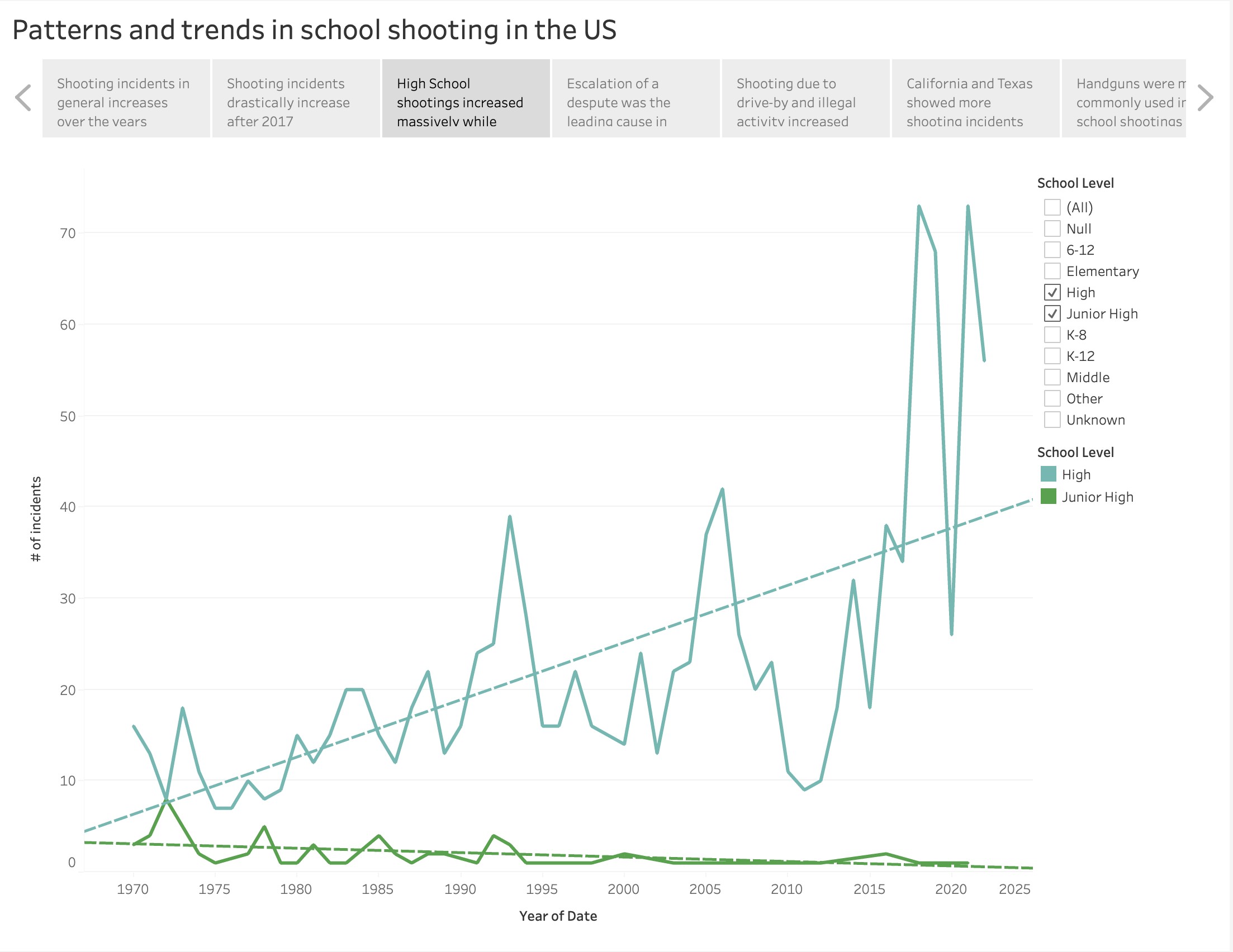


**Discovery 4**: Shooting incidents drastically increase around and after 2015.

The line chart also reveals there is a sudden rise shootings after 2015. This rise brings the number of shooting to record high numbers which far exceeds any of the previous numbers. This is true in case of most school levels. This information can be seen more clearly when all the lines in the chart are high- lighted after 2015. The trend lines generated show much steeper gradient for almost all school levels. Such instance of the chart is shown below.

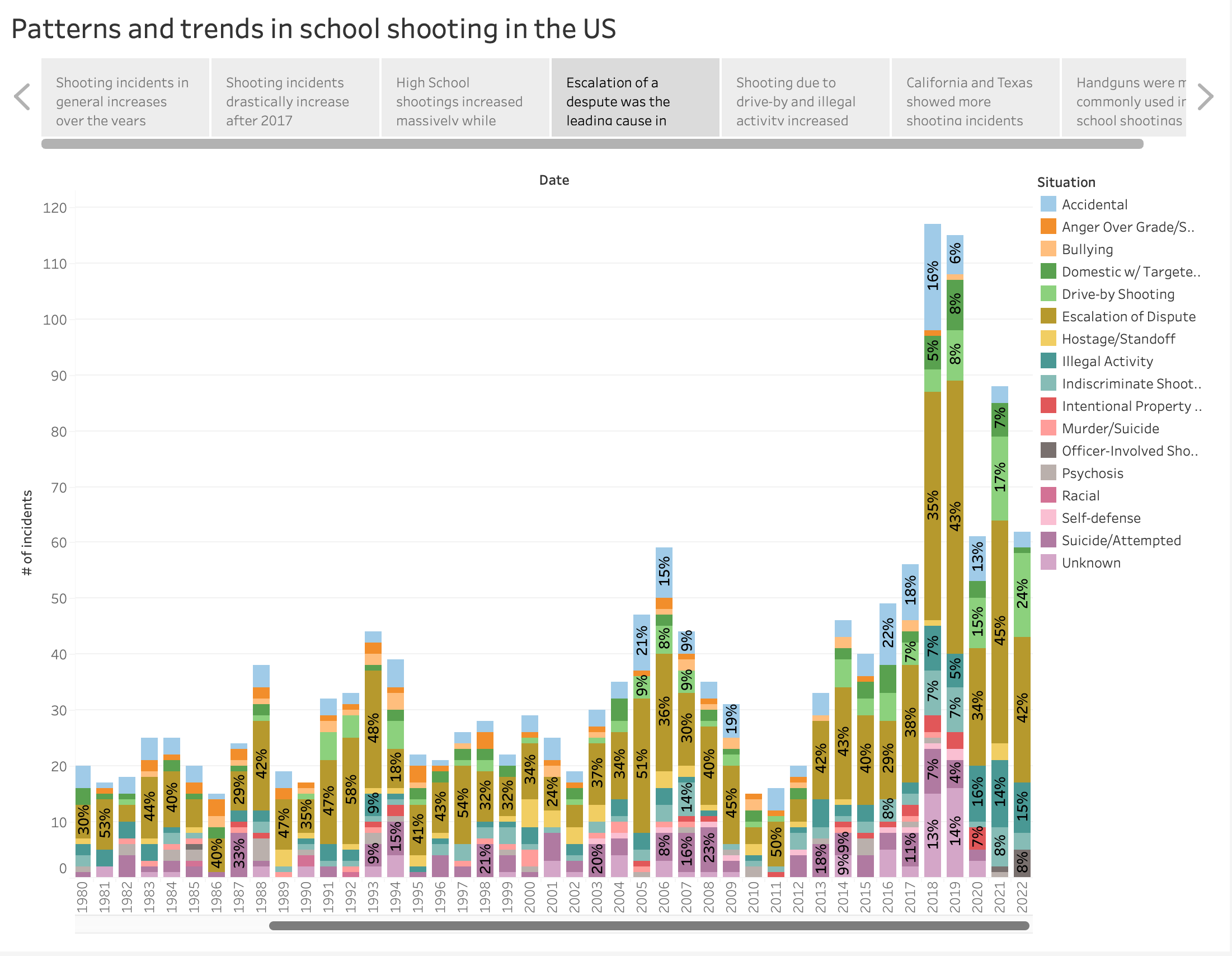


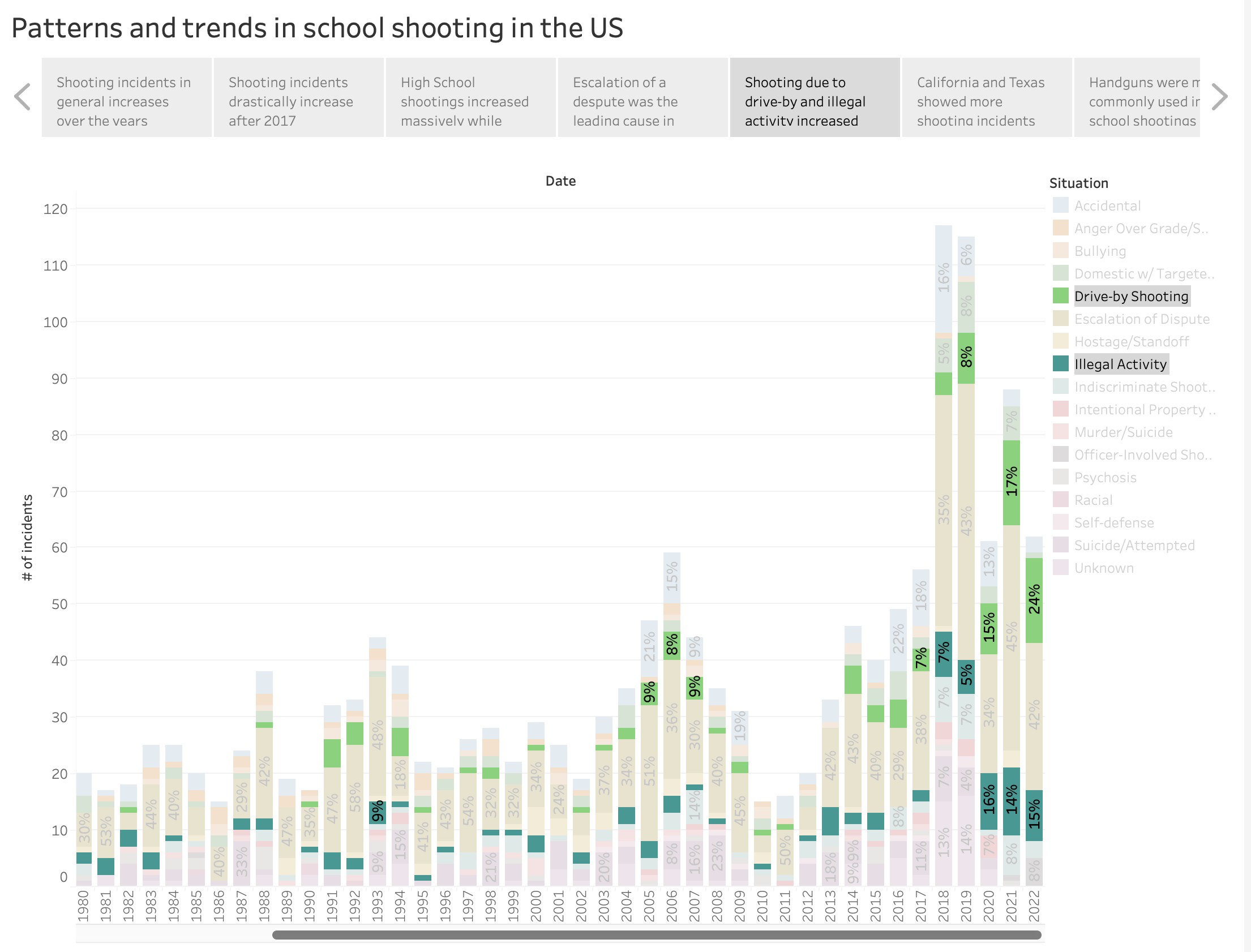
**Discovery 5**: High School shootings increased massively while junior high de- creased. High scool shooting occurences have almost always been higher than any other school levels, evident from its line staying higher than incident lines for other school levels. Furthermore, the school level seemed to have the highest increase in school shootings as described by the steepest gradient line. On the other hand, on careful observation, it is found that only junior high school level has had an overall decreasing shooting trend compared to other school levels. The difference in both the school levels can be seen clearly by selecting only the two levels from the School Level filter.



**Discovery 6**: Escalation of a dispute was the leading cause in almost all years.

Escalation of a dispute was the major factor in almost every year. Plot of the reasons for shooting over the years revealed that escalation of a dispute had much higher percentage in the stacked bars than any other reason individually. This was true for most years.



**Discovery 7**: Shooting due to drive-by and illegal activity increased after early 2000s Highlighting each factor from the filter indivually revealed that shooting due to illegal activity or driving by had an increase in taking part in school shootings after 2000s. The two factors were almost neglegible before the 2000s but have had a noteable increase afterwards.

# Reflective Discussion

This project focused on visualising School shooting data to uncover trends and information which could be beneficial to an interested data analyst. I believe that the project does succeed in revealing information which would be otherwise difficult when attempted without visualisation techniques. The work and the results obtained does stay true to the questions asked and helps to visualise information as clearly as possible within its set constraints. I think the research questions asked could be more specific, as the ideal and more elaborate solutions demanded by the questions would require creating more than one dashboards of visualisation. This module was my first time working in Tableau, and I firmly believe the software has widened my views on what can be achieved through visualisation. It is impressive how the software is able to provide such flexibility while staying user-friendly. I would find it more helpful if te software allowed more data-preprocessing power. Throughout the CS5803 module, I have understood the necessity of suitable visualisation techniques and the effectively few good rules of thumb which make the concepts easier in the long run. I, however, still find myself struggling with coming up with the most ideal solution to a visualisation problem, as I somewhat doubt the plots I think would be suitable for a problem, in some cases. I believe this is a struggle that can

effectively be solved through practice, and I aim on solving as much problems as I can to build my confidence in data visualisation.

# Conclusion

This project involved using various visualisation techniques to try to find trends in school shooting data. The K12 school shooting database is a large dataset for its domain, and offers potential for many insightful visualisations.

This project started out with two specific objectives or research questions. The first one involved looking for similarities and differences in school shooting across the different states of the US. The approach to solving this involved building map visualisations to gather an overall impression from the data. It revealed that there were two states, California and Texas, which behaved differently com- pared to the other states. the states have exceptionally high level of shooting instances. When trying to look if different type of guns used may have had such an effect, it was found that most of the states had a similar gun distribution and, additionally, handguns were the most go-to weapons in these instances. This might point to the fact that handguns could be more cheaper or easily accessible, which obviously requires more looking into. The high occurrences in California and Texas might be explained by their gun legislation laws.

The second question focused on the patterns in school shooting over the years. The question was approached using two charts, one which looked at how shoot- ing incidents varied over the years for different school levels and as a whole, and another one which looked at the reasons for school shooting over the years. The first chart revealed that school shooting increased over the span of last 30 years, and high school has been the major setting for such events. The second chart revealed that escalation of a dispute has been the major factor for a school shooting in most cases every year, while reasons such as drive by shooting and illegal activities seem to be emerging.

# References

CS5803 Week 18 Teaching Materials (2023). Lecture 2: Encoding for Percep- tion. Available at https://brightspace.brunel.ac.uk, accessed 10th April 2023.

CS5803 Week 19 Teaching Materials (2023). Lecture 3: Representation. Avail- able at https://brightspace.brunel.ac.uk, accessed 11th April 2023.

CS5803 Week 20 Teaching Materials (2023). Lecture 4: Presentation. Available at https://brightspace.brunel.ac.uk, accessed 11th April 2023.