INFO247 Final Report

US Mass Shootings

—

Aakriti Kaul

Rajasi Desai

Shaivya Rastogi

Table of Contents

Project Goals

On February 14th, 2018 news broke that 17 students had been fatally wounded by a young gunman in **Parkland**, Florida[1]. Only 5 months prior to that, the horrific **Las Vegas** shooting[2] claimed over 50 lives. The nightclub shooting[3] in **Orlando** too flashes before our eyes when we begin this conversation.

With heated discussions, multiple global protests and activist movements encouraging a dialogue about the increase in gun violence, possible gun law reforms and the safety of citizens across the world, mass shootings[4] have been a headline for a long time now, especially in the United States. The recent shooting at the YouTube Headquarters[5] in San Bruno, propelled our group to understand the progression of mass shootings in the United States and to highlight these attacks on humanity which have become so commonplace and how they have developed across time - from the 60s to the current decade.

With the help of an extensive dataset, our project aims to shed light on the trends in US mass shootings over the past 6 decades, highlight the number of casualties, investigate the significance of attributes such as the year of an incident, social environment at that time, racial background of the perpetrator and other related data points.

---

[1] https://www.vox.com/policy-and-politics/2018/2/14/17013596/parkland-florida-high-school-shooting
[2] https://www.vox.com/2017/10/2/16399350/las-vegas-shooting-stephen-paddock
[3] https://www.vox.com/2016/6/12/11911616/pulse-gay-nightclub-shooting-in-orlando-what-we-know
[4] According to the Gun Violence Archive, a mass shooting is described as four or more individuals being shot or killed in the same general time and location
[5] https://www.vox.com/2018/4/3/17194258/youtube-san-bruno-california-shooting

Related work

Aakriti Kaul

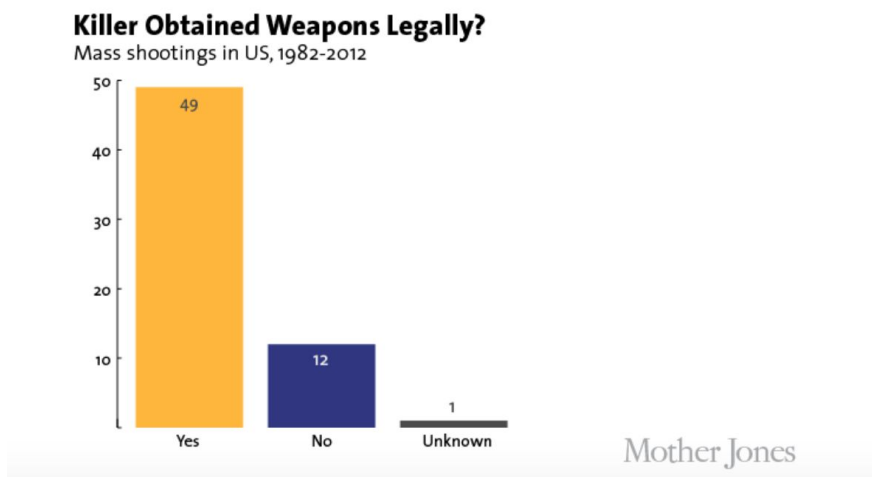1. https://www.motherjones.com/politics/2012/07/mass-shootings-map/

As part of the initial literature survey and investigative research, Mother Jones was cited as the most comprehensive list of incidents for over four decades. Being able to read the journalistic approach to this data helped us shape our storyline for the website.

- **We included a handful of cases also known as "spree killings"**—cases in which the killings occurred in more than one location, but still over a short period of time, that otherwise fit the above criteria.
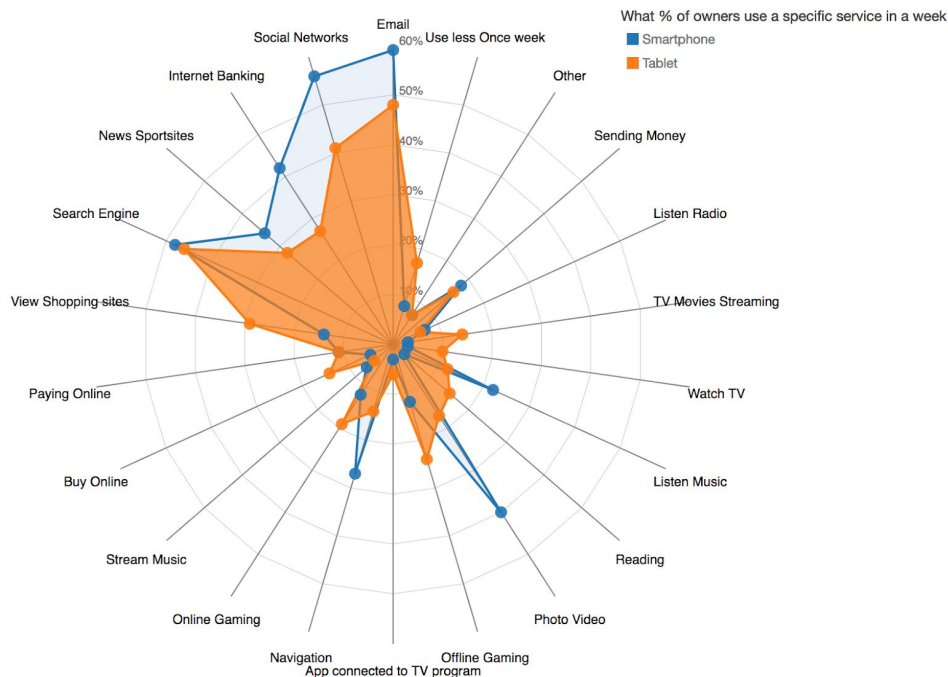
For more on the thinking behind our criteria, see these two explanatory pieces. Plus: more on the crucial mental illness factor, and on a barrage of state laws rolling back gun restrictions across the US. And: Explore the full data set behind our investigation.

Here are two charts detailing the killers' weapons:

Killer Obtained Weapons Legally?
Mass shootings in US, 1982-2012

2. http://bl.ocks.org/nbremer/6506614

A very vital chunk of our D3 visualization on interactive radar graphs is attributed to Nadieh Bremer's work shared on bl.ocks.org which was a website introduced to us through the i247 course. A variation to this was also incorporated within our website.

Figure: Radar chart titled "What % of owners use a specific service in a week" with legend — Smartphone (blue), Tablet (orange). Categories: Email, Social Networks, Internet Banking, News Sportsites, Search Engine, View Shopping sites, Paying Online, Buy Online, Stream Music, Online Gaming, Navigation, App connected to TV program, Offline Gaming, Photo Video, Reading, Listen Music, Watch TV, TV Movies Streaming, Listen Radio, Sending Money, Other, Use less Once week. Radial axis: 60%, 50%, 40%, 30%, 20%, 10%.

3. https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/?utm_term=.ac87488b532d

One of the most impactful resources that we came across was this Washington Post article that allowed us to understand how a narrative can flow with text accompanied visualizations.

1,081 killed

The people who were killed came from nearly every imaginable race, religion and socioeconomic background. Their ages range from the unborn to the elderly; **176 were children and teenagers**. In addition, thousands of survivors were left with devastating injuries, shattered families and psychological scars.

CLICK ON AN ICON FOR DETAILS ABOUT EACH VICTIM.

Rajasi Desai

1. https://www.usnews.com/news/best-states/florida/articles/2018-03-15/timeline-of-pivotal-moments-in-us-gun-control-history

This article gives a timeline of the gun control laws in the US since 1791. I used data from this article along with this article to create the Gun Control Timeline.

2. https://edition.cnn.com/2017/10/03/americas/us-gun-statistics/index.htmlhttps://edition.cnn.com/2017/10/03/americas/us-gun-statistics/index.html

This article was extremely useful in terms of statistics, which were used on the website as well as during our presentation in the project showcase. Details such as there are 89 guns per 100 people in America are powerful in itself, but adding details of the second spot, Yemen, with 55 guns per 100 people was more impactful during the presentation.

Americans own more guns per capita than residents of any other country

COUNTRIES WITH THE MOST GUNS Average firearms per 100 people		COUNTRIES WITH THE FEWEST GUNS	
USA	89	0.1	Tunisia
Yemen	55	0.3	East Timor
Switzerland	46	0.4	Solomon Islands
Finland	45	0.4	Ghana
Serbia	38	0.4	Ethiopia
Cyprus	36	0.5	Singapore
Saudi Arabia	35	0.5	Indonesia
Iraq	34	0.5	Eritrea
Uruguay	32	0.5	Fiji
Sweden	32	0.5	Bangladesh

CNN   Source: Small Arms Survey, 2007

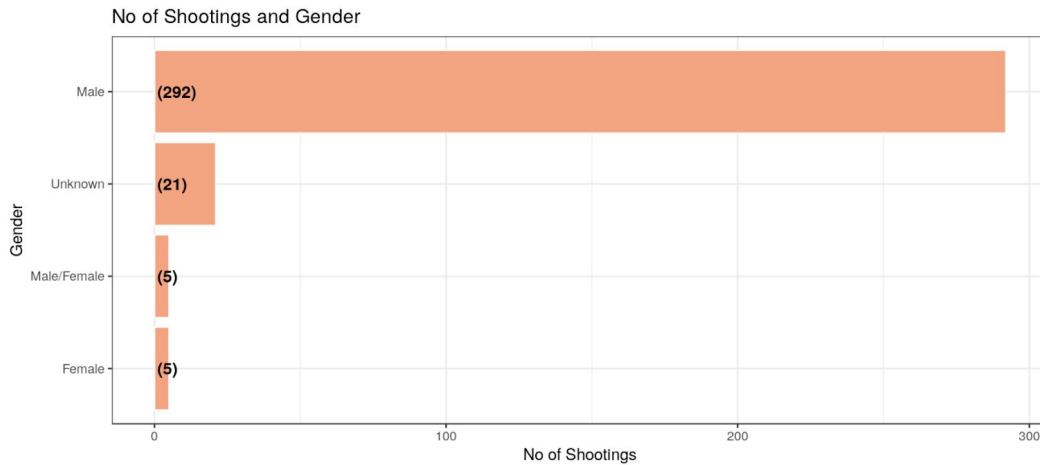3. https://www.kaggle.com/ambarish/us-mass-shootings-analysis

This analysis was helpful in terms of narrowing my focus on gender and race distribution of shooters. Graphs 11 and 13 show race and gender distributions.

11 Shootings per Gender

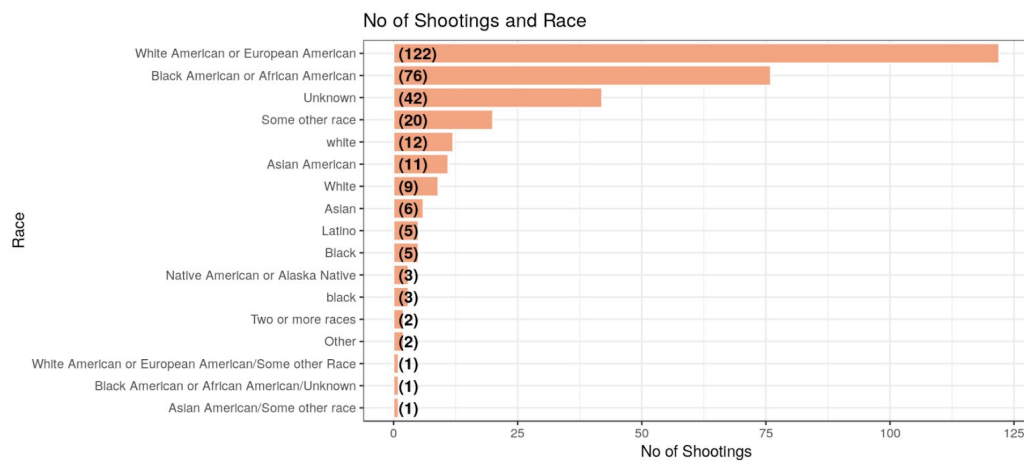Males are engaged far more shootings than females

Code

No of Shootings and Gender

13 Shootings and Race

White American shooters are more common than the Black American shooters

Code

No of Shootings and Race

Although a simple bar graph was sufficient in representing the race distribution of the shooter, we thought a tornado chart (or population pyramid) was a better way to represent the gender of the shooter.
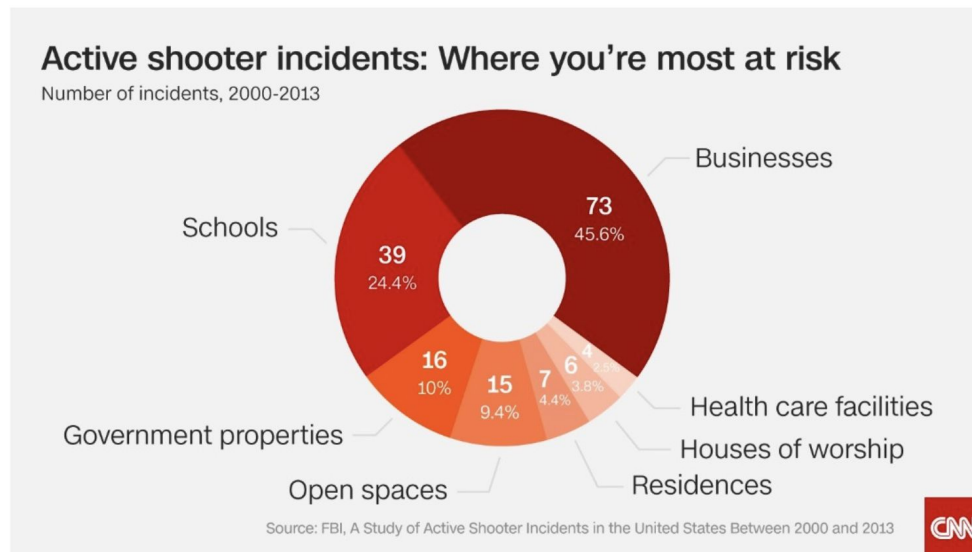
Shaivya Rastogi

1. https://www.cnn.com/2016/06/13/health/mass-shootings-in-america-in-charts-and-graphs-trnd/index.html

This CNN article contains 9 charts that show how mass shootings in America are a serious problem. We really liked the idea of showing the geographic distribution of the number of shootings via a choropleth, and normalizing it by showing the incidents per 5 million people. The chart looks neat, and clearly shows which areas have higher number of shootings per 5 million people. Inspired by this, we decided to show state-wise distribution of incidents in a similar manner using a choropleth map.
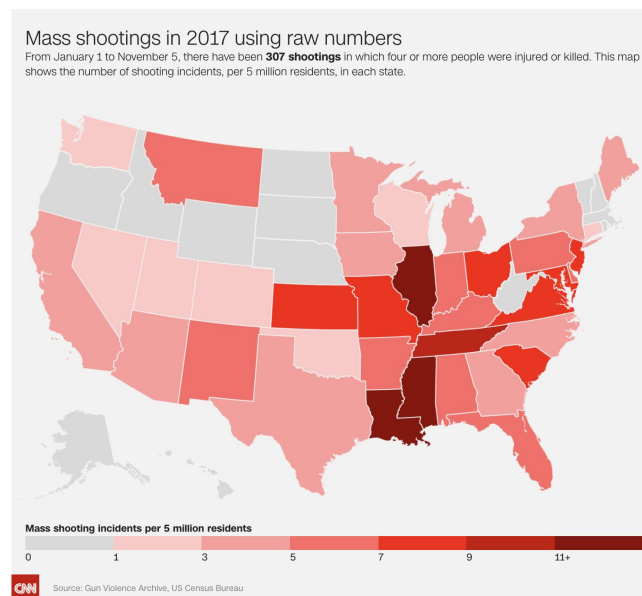
Another chart that we really liked showed the distribution of incident locations in a donut chart. This chart shows that schools and businesses are the most popular target

locations. We decided to include a similar visualization in our website in the form of an interactive radar chart.

Active shooter incidents: Where you're most at risk
Number of incidents, 2000-2013

2. http://www.storybench.org/mass-shooting-data-visualizations/

The Storybench visualization shows a number of aspects of mass shootings. It helped serve as an initial literature reference for diving into the combination of gun laws and mass shootings. One of the visualizations shows how mass shootings have been occurring more frequently for the last few years, hence gave us the idea to start by showing the increase in number of incidents across decades. Since it shows visualizations in multiple colors, it also helped us think of our color scheme - black and red.

3. https://www.vox.com/policy-and-politics/2018/2/21/17028930/gun-violence-us-statistics-charts

Another article/visualization we referred to is this Vox article on gun violence. Similar to the second reference I mentioned, it became part of our literature survey on seeing how

big a problem guns is in America, and how gun control has worked in other countries has worked in the past. We ended up including a statistic related to this topic as a result of this article. We had 'Mental Health' as one of the features in our dataset, and we were already skeptical on whether to use it in our visualizations or not. This article also shows how people with mental illnesses are more likely to be victims of gun violence, and not perpetrators, and helped us decide against using it in our website.

Preliminary Work And Progress Steps

1. Literature survey and exploration of existing data journalism articles to understand the product of an unbiased take on the topic for a holistic end narrative
2. Identify and collect suitable data to incorporate statistics that may not be readily available through the main dataset
3. Processing and cleaning the dataset to allow ease of integration into Tableau and D3 implementations. While all original variables were retained, processing included feature generation used in analysis such as incident areas, decade and normalized numerical values
4. Exploratory Data Analysis using Tableau to identify possible relationships between variables and also understand the distribution of null values to effectively tackle these during visualization
5. Paper prototyping for the graphs and website layout (See here for initial sketches)
6. Peer review and feedback on initial sketches and the feedback from the mid-project presentation allowed us to identify gaps in the narrative and make changes to ideas where deemed necessary. For instance, a hypothesis to study the relation of mental health in perpetrators was dropped after feedback from the live presentation
7. Setting up the website template and Github repository to begin project collaboration

8. Development of visualizations and the website narrative content with each team member owning a specific part and providing feedback on the rest through heuristic evaluation
9. We conducted usability tests with participants from within and outside the MIMS program to avoid bias
10. Feedback provided by peers, instructors and external visitors during the project showcase was incorporated to allow a holistic visualization project
11. During the project showcase, we received more feedback from peers as well as outside visitors regarding the visualizations. We were also able to observe how easy or difficult it was for people to understand the intended message
12. Documentation of progress and final consolidated report was completed by all team members collectively

Description of Visualizations

FATALITIES/INJURIES ACROSS YEARS

The first thing we wanted to show was how the number of mass shooting incidents have changed over the years. Through our EDA, we found that the number of incidents, fatalities, and injuries have been increasing over the decades. We normalized the numbers according to the population in each year to incorporate the impact of changing population. Despite that, we see an overall increase. We see a slight dip in the end, but that's because 2018 has only a few months of data.

To show this trend, we decided to make a **line graph in Tableau**, which makes it very easy for the user to read and identify trends. It also enabled us to show the number of fatalities, incidents, and injuries in one graph. To maintain consistency through the website, we decided to depict Fatalities using red color.
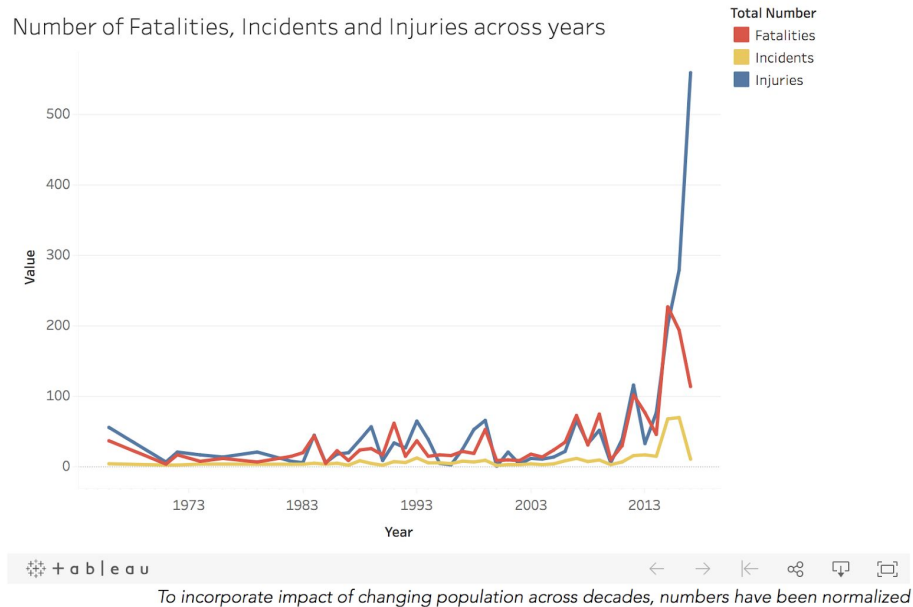
Number of Fatalities, Incidents and Injuries across years

To incorporate impact of changing population across decades, numbers have been normalized

Figure 1. Fatalities, Incidents and Injuries across years

GEOGRAPHIC DISTRIBUTION OF INCIDENTS

After showing the increase in number of shootings over the last few decades, we decided to show the state-wise distribution of the incidents. To achieve this, we had to first extract the state code for all the shooting locations. After doing that, we decided to show three parameters in this distribution - number of mass shooting incidents, number of fatalities, and number of victims (fatalities plus injuries). We ended up showing this through a **choropleth map in Tableau**, with the three parameters provided as filters for the user to choose. The states are shaded according to one of these three parameters selected by the user. The map aids users in identifying trends, if any, across certain regions of the country.

Like in the first visualization, we normalized the numbers to account for the changing population. Since the states have different populations and different sizes, showing the raw numbers of incidents, fatalities, or victims could be misleading to read, as larger and

more populous states could appear to have more number of incidents just because of their size. To ensure we don't mislead the users, instead of displaying raw numbers, we showed the numbers per 5 million residents. Hence, the values are normalized to take care of both, changing population, and different state populations.

We kept the color of the choropleth to be shades of red to maintain consistency in the website. Darker shades correspond to higher numbers, keeping it consistent with what a user perceives in the first look.
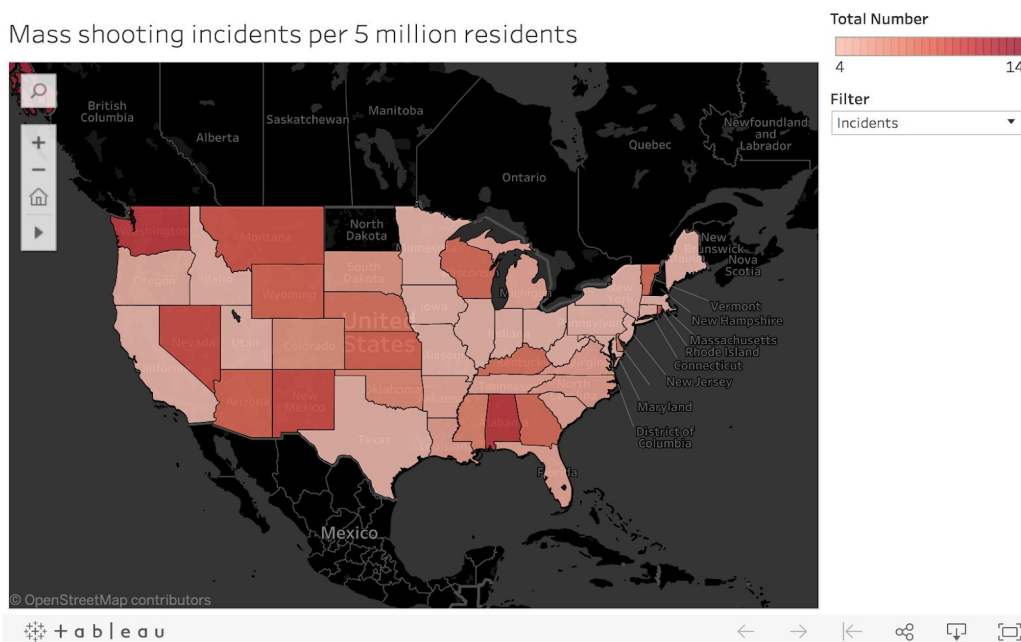
Figure 2. State-wise distribution of number of incidents

Looking at Figure 2, we can see easily see that Washington, Alabama, Nevada and New Mexico have a high number of incidents per 5 million residents. In general, central US seems to have a relatively high number of incidents.

When viewing the number of victims, as in Figure 3, we see that Nevada has a very high number of victims as compared to other states. This visualization doesn't help explain

much since all the other states are colored in a similar shade because of the difference in the number scale.
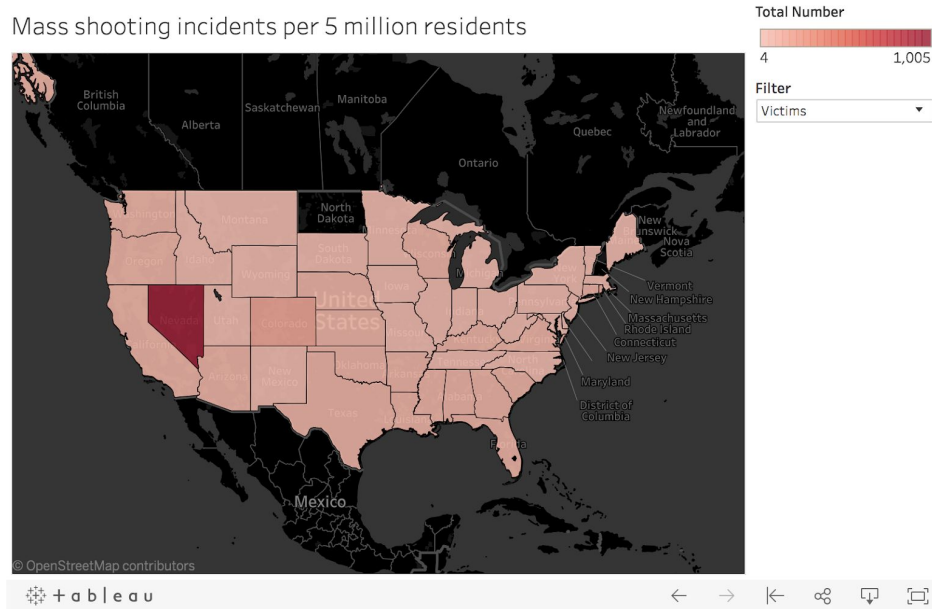
Figure 3. State-wise distribution of number of victims

To be able to look at a more descriptive visualization for other states, one option is for the user to exclude Nevada from the map. Tableau allows users to select/exclude particular data points. If we exclude Nevada, the visualization looks more informative, as shown in Figure 4.
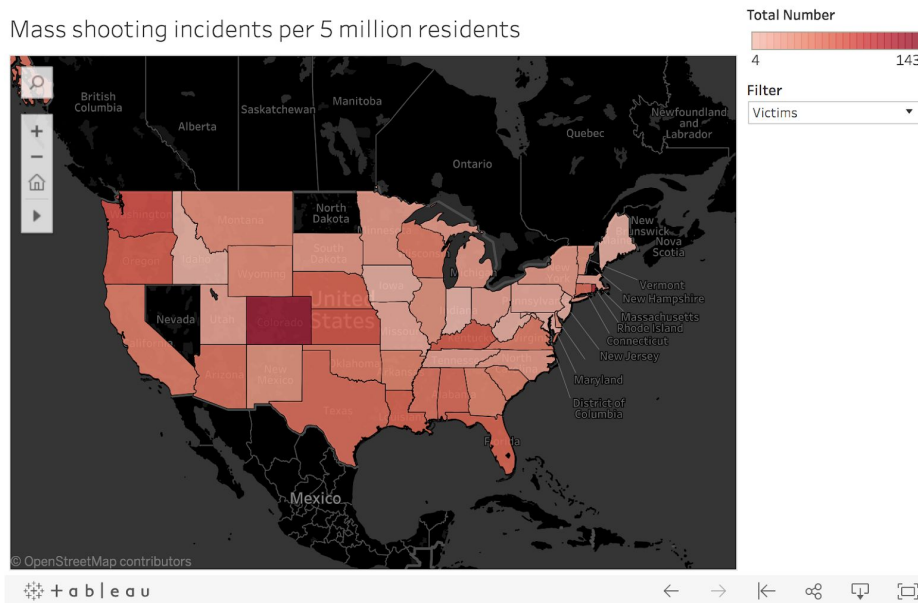
Figure 4. State-wise distribution of number of victims (Excluding Nevada)

INCIDENT AREAS ACROSS DECADES

16 weeks into 2018, CNN reported an average of 1 shooting per week in schools. In the aftermath of the Parkland shooting in Florida, there have been multiple global protests and activist movements for the safety of students. While school shootings seem to be at a peak, we decided to delve deeper into the distribution of mass shooting incidents across the years based on the type of location, with data categorized into 9 major incident areas which included:

- Nightclubs

- Open Public Areas

- Places of Worship

- Schools

- Residential Areas

- Restaurants

- Workplace

- Stores and Malls

- Others

Leveraging work by Nadieh Bremer and Chris Zhou, we customized an interactive D3 visualization in the form of a spider chart with each axis denoting one of the nine incident areas and a shaded polygon spread across the scale to show the number of victims, fatalities and injuries per decade - ranging from the 1960s to 2010s. As all three of these variables are closely intertwined, we wanted to see any correlation or pattern if possible.

Given that each group - victims, fatalities, injuries - overlapped at certain points visually in the default collapsed view, a user can also expand the visualization into a faceted spider graph using a checkbox that splits the graph across the three groupings. The choice of a spider graph was to be able to compare aggregated numbers across locations, groupings as well as decades in a clear manner that did not simplify the information.

While the "web" does not show the scale, hovering over each point of intersection provides users with the dynamically updated distribution of values. The scale values change as each decade changes. As with all data used in this project, numbers have been normalized to account for the impact of changing population across decades.
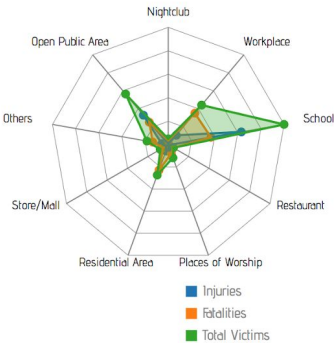
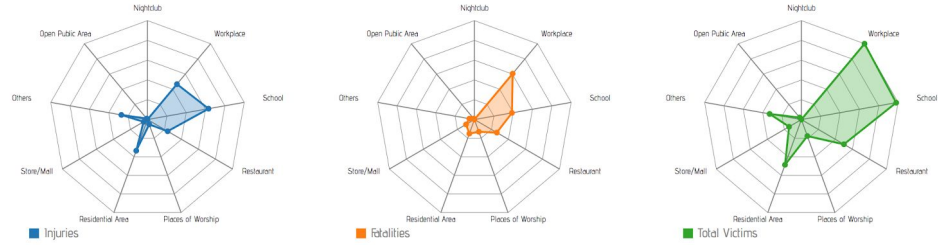Figure 6. Default collapsed view for incident area radar chart

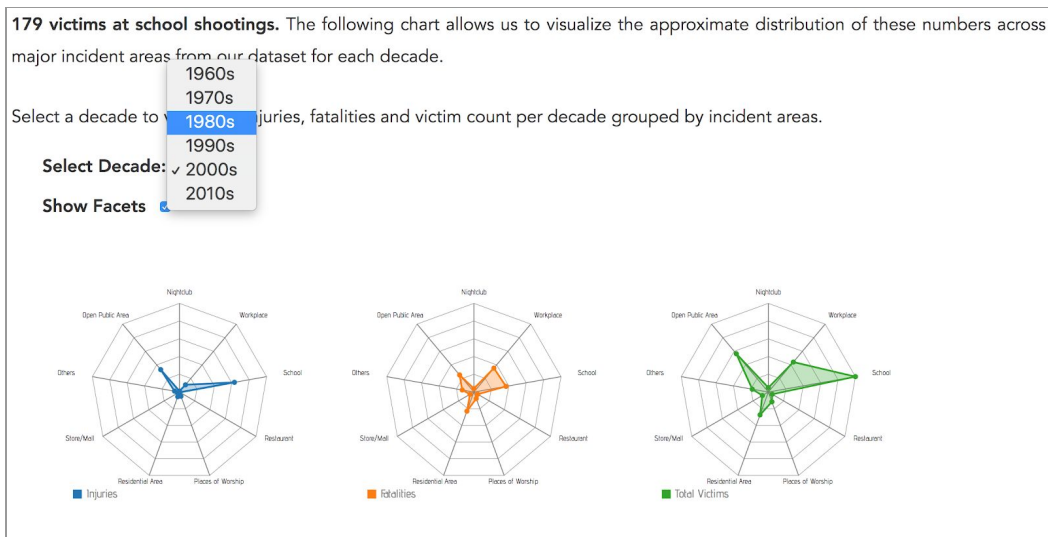Figure 5. Facet view for incident area radar chart

Figure 7. Decade selection for incident area radar chart

RACE

Often, there are preconceived notions about the race of perpetrators in violent crimes. We were interested in exploring the distribution of race of the shooters for the mass shootings. The following bar chart shows that more than 45% were Whites followed by African Americans at 26%. This was followed for Asian Americans and Latinos. It is worth mentioning that 22% of the values here are unknown and knowing these values might have changed the distribution of the graph.
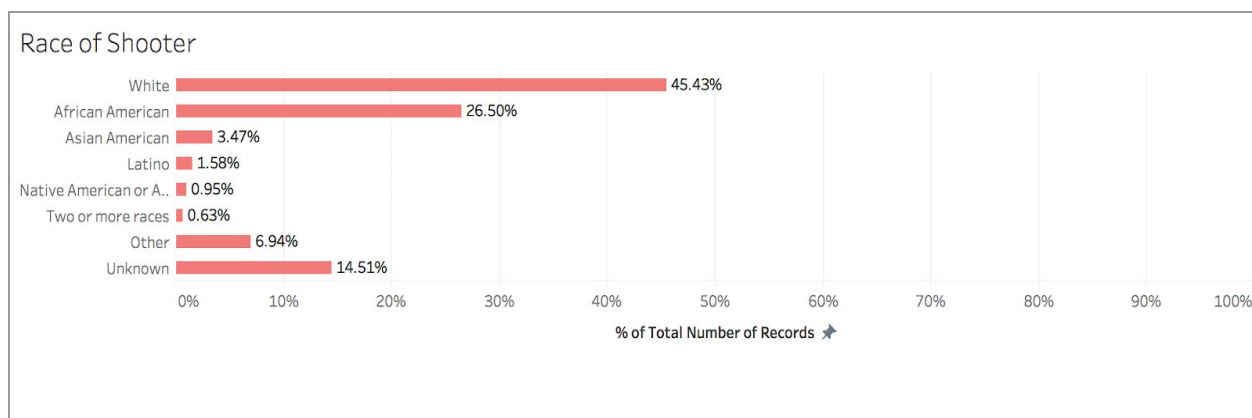
Figure 8. A numerical comparison of shooter race

After looking at the race distribution of the shooters, the next step was drilling down on the gender distribution of the shooters.

GENDER

With the recent shootings at YouTube carried out by a female perpetrator, we were curious to know about the distribution of male and female perpetrators in past shootings. Although we hypothesized the number of male perpetrators to be higher, we were surprised by the humongous difference in numbers. Out of __ shooters, the distribution stands at **292 males** and **5 females**. The best way to portray this information, in our opinion, was through a tornado chart (aka population pyramid). Since this chart could not be made directly on Tableau, we plotted individual graphs for men and women perpetrators and combined them on Adobe Illustrator.

Incidents with shooters of both genders and unknown genders have not been considered in this graph.
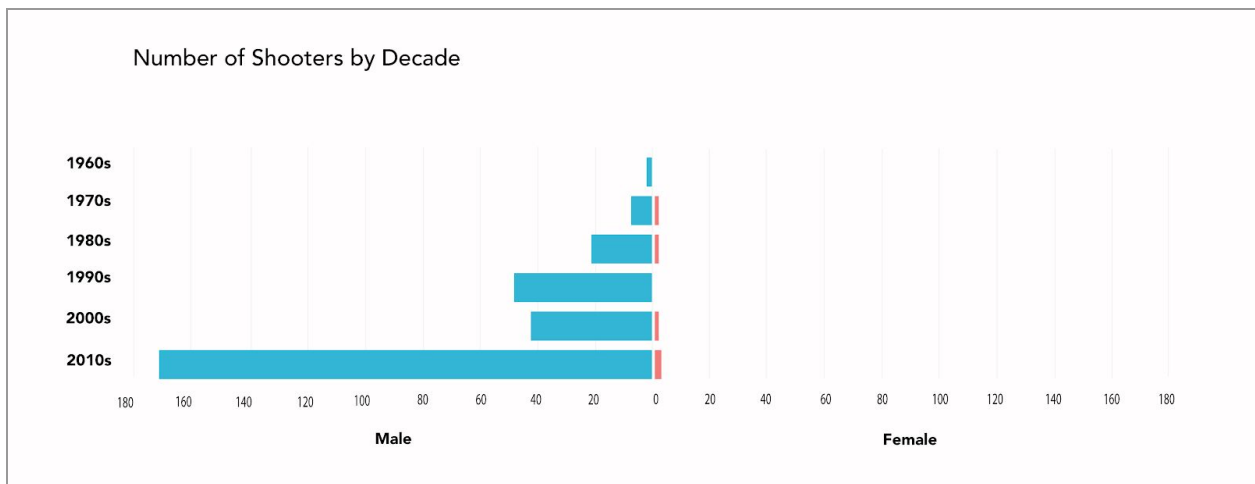

Figure 9. Population pyramid showing the drastic difference in genders of mass shooters

GUN-CONTROL TIMELINE

Numerous attempts have been made to tighten gun-control in the United States. While some attempts have been successful, some proposed laws have not seen the light of day. With this visualization, we attempt to depict milestone gun control acts in American History, the political and social incidents which inspired these acts as well as proposed laws which did not pass at a national level.

The aim of this timeline is to highlight the influence of politicians and ruling parties on the gun control acts passed as well as shed light on shootings which were instrumental in passing(or proposing) of said gun control acts. For example, tragic incidents like the heartless assassinations of President John F. Kennedy (1963) in Texas, former Attorney General and Presidential Candidate Robert Kennedy(1968) in Los Angeles, Civil Rights Activist Martin Luther King Jr (1968) in Memphis, led to President Lyndon B Johnson pushing for **The Gun Control Act** (GCA) of 1968. Similarly, the **Brady Handgun Violence Protection Act** was passed in the name of the White House press secretary who was left permanently disabled after the attempted assassination of then President, Ronald Reagan.

Apart from political incidents, social activism has also resulted in positive changes in gun laws. After the recent shootings in Parkland, Florida on February 14, 2018, the national uproar resulted in strikes and marches carried out by students as well as parents. As a result of this public outcry, two laws were passed - one on a state level and the other on a Federal level. The **Marjory Stoneman Douglas High School Public Safety Act** raised the minimum age for buying rifles to 21 and introduced tighter background checks while the **STOP School Violence Act** was signed into law which increased funding for metal detectors and other safety measures. This goes to show that power of politics and voice of the people has the potential to revolutionize gun laws in the United States.
It will be interesting to see the fall or rise of gun related violence after each law was passed and this can be an area of further exploration for this topic.
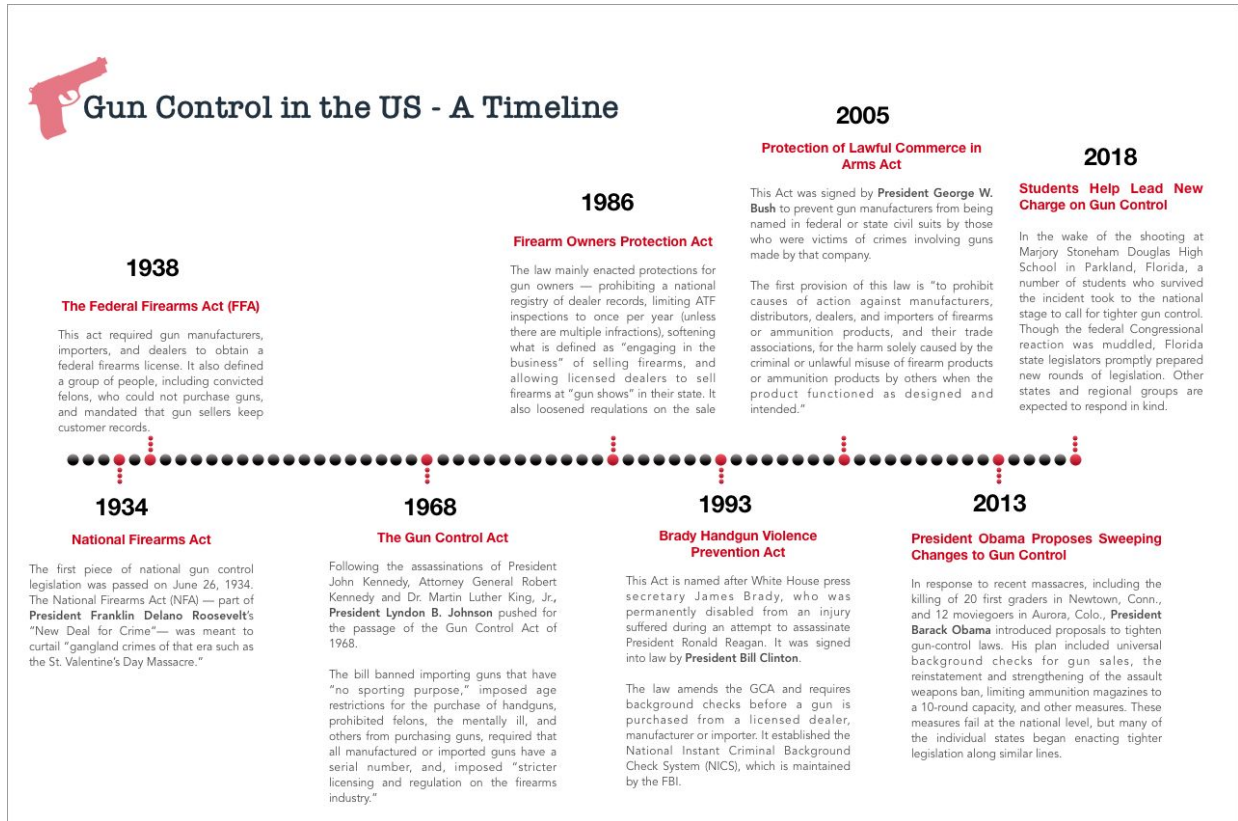
Figure 10. Infographic depicting political influence on incidents

Data

The comprehensive dataset provided on Kaggle US Mass Shootings was selected as the primary data source for our project. Over the course of the project development, we incorporated qualitative information from the source for FiveThirtyEight's story on Gun Deaths In America and also standalone statistics found from reports attributed here.

In order to account for the changing population across years and to normalize the choropleth visualization for the population of states, we used external data sources including this and this.

Tools

The dataset we used had a few missing values, hence we first needed to explore the dataset to understand the missing values and how to deal with them. We used **MS Excel** for the first level of analysis and to understand the preprocessing that we needed to carry out. Once we understood the data and took care of the missing values, we then used **Tableau** to perform Exploratory Data Analysis to gain some interesting insights and identify trends in our data.

After EDA, we used pen and paper to draw initial sketches of the visualizations we planned to have in our website and for initial user testing. We also sketched out the layout of the website we intended to create. We decided to use **Tableau** for a few of our visualizations. We used **D3.js** for creating the interactive visualization about incident locations. We used **Sketch** to draw the Gun Law timeline and to edit isotypes for the statistics we presented.

In order to create the website we used to present our visualizations, we used **HTML**, **CSS**, and **Javascript**. We also used GitHub for collaborating and sharing our code.

Results

We carried out heuristic evaluation and usability tests to evaluate the usability of the website.

HEURISTIC EVALUATION

The website was evaluated on the basis on Nielsen-Molich Principles. Three to five evaluators are enough to carry out an efficient heuristic evaluation. Since our team consisted of three members, each person carried out the evaluation individually and

reported their findings to other members. The errors were fixed and a final round of evaluation was carried out on April 30, 2018.

The heuristic evaluation revealed the following flaws:

Match between system and real world

We observed that the website did not follow a logical order. The gun control timeline was oddly placed and therefore did not lead to a cohesive flow in the website. We reordered the website to shift the gun control section at the end of the website.

Consistency and Standards

1. The website had inconsistent font size and styles which was caught during heuristic evaluation and was fixed.
2. The choropleth for statewide distribution of incidents, victims and fatalities was initially blue in color, which was not consistent with the black and red color scheme of the website. Therefore, the color of the map was changed to red, with a black background.

Before:
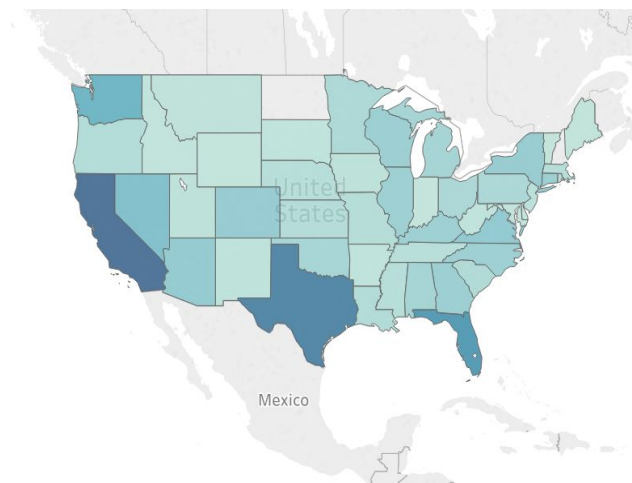
After:

Mass shooting incidents per 5 million residents
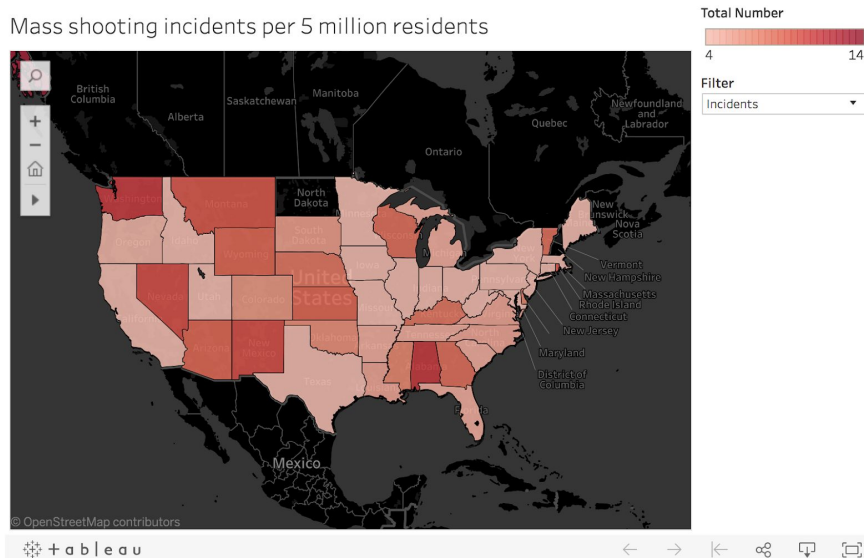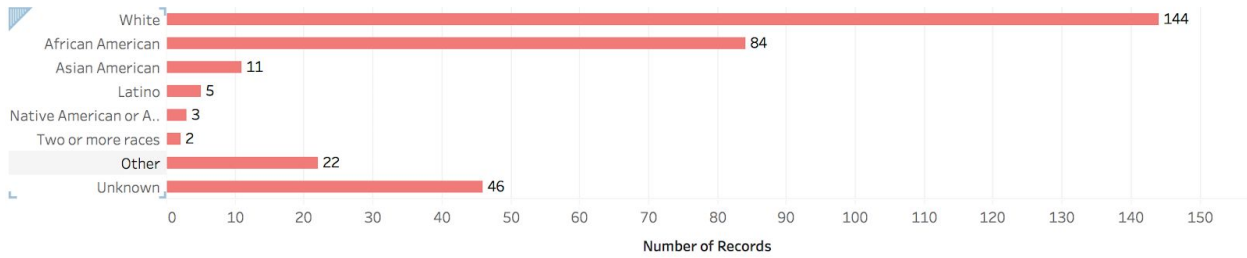
Aesthetic and minimalist design

The website was text heavy, which caused clutter on the website. Some of the information was unnecessary and did not add value to the website. This information was removed, thus keeping a minimalist design.

Recognition rather than recall

The graph with race of the shooters revealed numbers only on hovering. The graph was represented as numbers but we realized this was not giving users a clear understanding of the percentage of users of a particular race. In-order to reduce cognitive overload for users, we changed the graph to add percentage values and added values for the graphs next to the bars.

Before:

Race of Shooter

White — 144
African American — 84
Asian American — 11
Latino — 5
Native American or A.. — 3
Two or more races — 2
Other — 22
Unknown — 46

Number of Records

After:

Race of Shooter

White — 45.43%
African American — 26.50%
Asian American — 3.47%
Latino — 1.58%
Native American or A.. — 0.95%
Two or more races — 0.63%
Other — 6.94%
Unknown — 14.51%

% of Total Number of Records
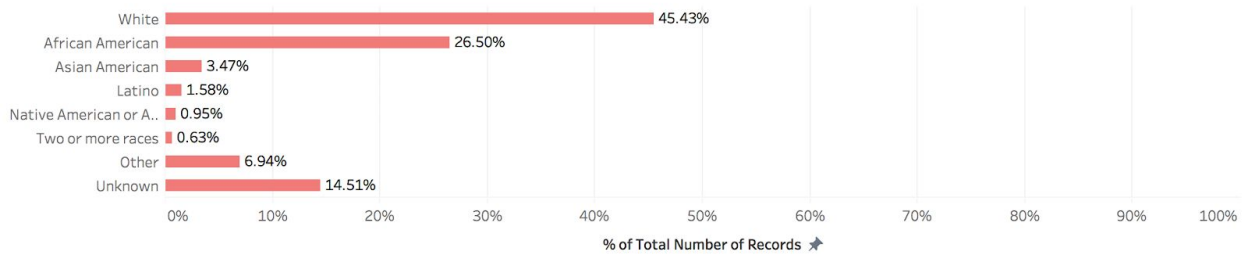
USABILITY TEST

We ran usability test on **five** people in our common circles. Since the website was targeted towards general public, we ensured that the participants were not restricted to our MIMS cohort.

We discovered the following usability issues from our tests:

1. The sources for the statistics were not mentioned. This was pointed out by one user and it was fixed immediately - the source was named and a link was provided to the source.

The US makes up less than 5% of the world's population, but holds 31% of global mass shooters.
Source: CNN

The number of civilian-owned firearms in the US is 88.8 guns per 100 people.
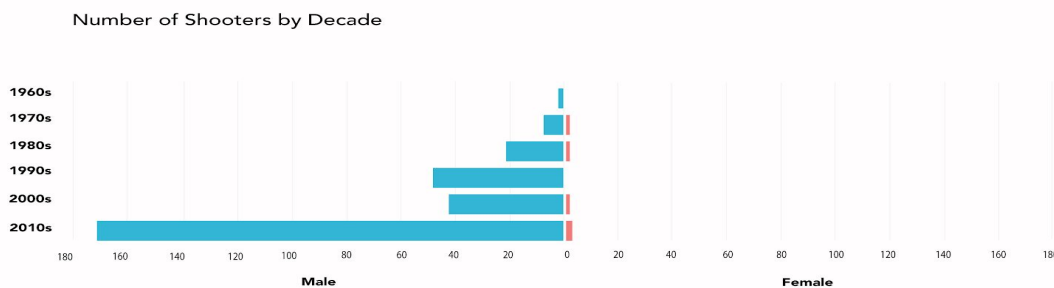Source: CNN

The cost of US gun violence is almost $2.8 billion a year.
Source: Quartz

2. The number of shooters by decade graph was first made in Tableau as a dashboard, but 2/5 users were confused as the graph had two '0' points on the X axis. Therefore, the graph was edited on Adobe Illustrator with one '0' point.

Before:
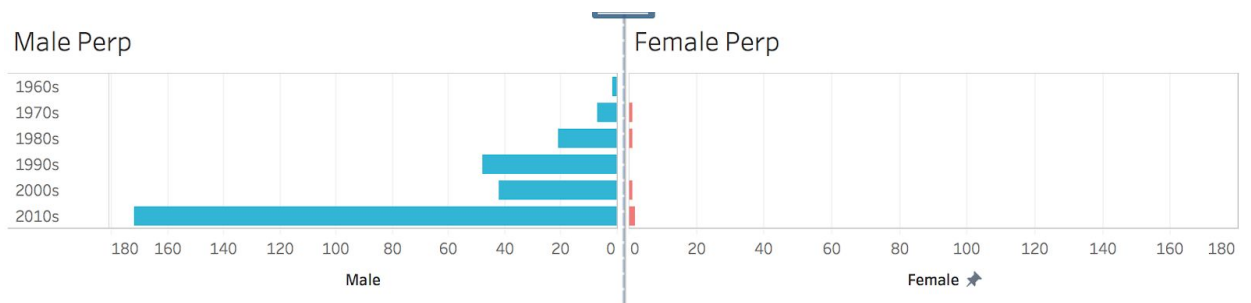
After:

3. The gun control timeline was text heavy and we wanted to ensure it is easily readable. 4/5 users could not name the key contributors in gun-control acts while 2/5 users thought the font was overpowering. In order to account for these issues, we highlighted the key players in bold and changed the color of the text from

black to grey. In order to confirm that the change of text color improved the usability, we showed all 5 users the same timeline with the same details except for the font color. ⅘ users preferred the grey text over black.

Before: (Black font color and no emphasis on key players)
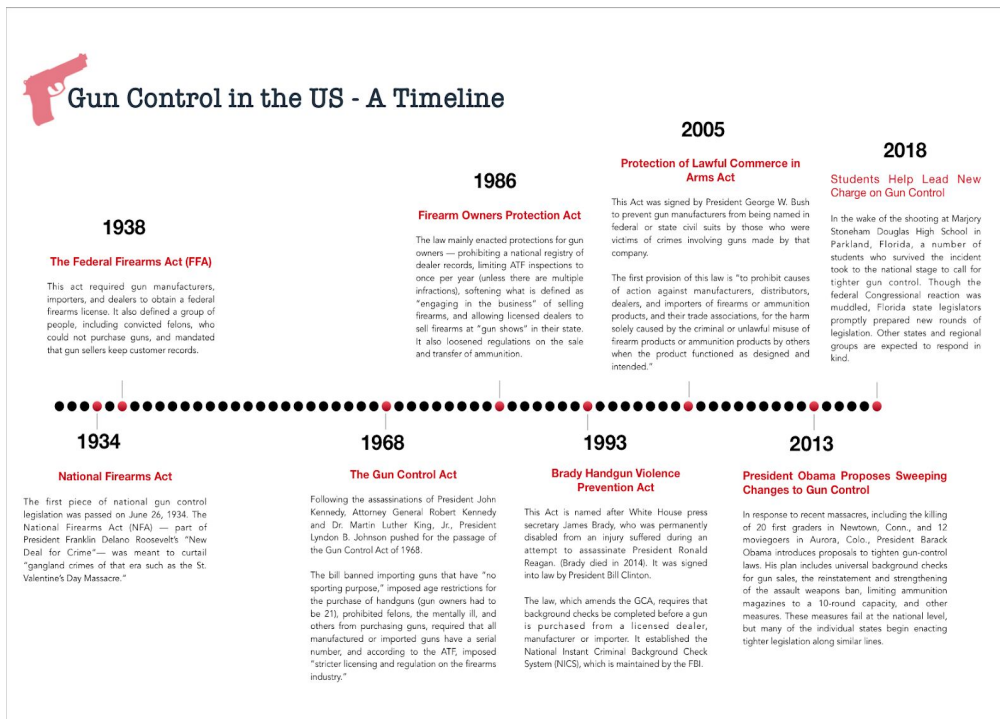
After: (Font color changed to gray and key players highlighted on bold)
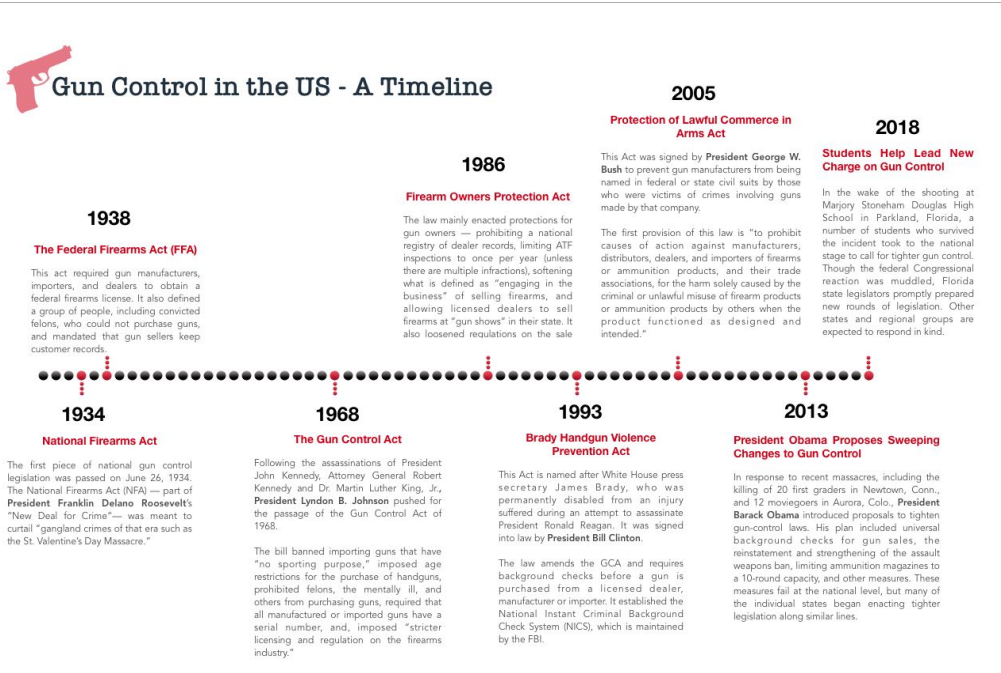
EXPERT EVALUATION

Our visualization was evaluated by an expert, Prof. Marti Hearst, who gave us great insight on the positives and more importantly, areas of improvement for our final product. Her feedback included:

1. The choropleth of the statewide distribution of victims, fatalities and incidents was not normalized for population of the states. Therefore, states with high population showed high number of incidents. This feedback helped us understand the limitations of the current implementation of the graph and therefore, we modified the choropleth to show incidents, victims and fatalities per 5 million residents.

2. The slider for the choropleth was stagnant with a maximum value of 36. This should have ideally changed with every category. Prof. Hearst pointed this out and it was immediately fixed.

3. Prof. Hearst suggested the website would look more pleasing to the eye with a few aesthetic changes to the layout. With this in mind, we increased the font size

to make the text more legible, added headings and modified graph alignments to enhance the look of the website and make it easier to read.

Links

To see our final project, please click here.

Code is publicly available here.

Contribution

Task	Proportion of Workload		
	Rajasi Desai	Aakriti Kaul	Shaivya Rastogi
Project Source Ideation	33.33%	33.33%	33.33%
Data Collection	33.33%	33.33%	33.33%
Data Cleaning and Processing	33.33%	33.33%	33.33%
EDA	33.33%	33.33%	33.33%
Visualization: Number of Fatalities/Injuries across years	-	-	100%
Visualization: State-wise distribution of incidents, fatalities, victims	-	-	100%
Interactive visualization: Incident locations across decades	-	100%	-

Visualization: Distribution of race of shooter	100%	-	-
Visualization: Distribution of gender of shooter	100%	-	-
Infographic: Gun law timeline	100%	-	-
Finding statistics	50%	50%	-
Color and Styling	33.33%	33.33%	33.33%
Website Text	50%	50%	-
Website Creation	-	-	100%
User Testing	33.33%	33.33%	33.33%
Final Report	33.33%	33.33%	33.33%

References

HEADER IMAGE SOURCE:

Billboard

WEBSITE TEMPLATE SOURCE:

Template

CODE SOURCES:

Nadieh Bremer's Radar chart

Chris Zhou's interactive visualization

DATA SOURCES:

Gun Deaths In America

Kaggle: US Mass Shootings

Appendix

INITIAL SKETCHES: