- **Basic Info**
  Title   **:** US Mass Shootings (1966-2017)
  Team   **:**
      Junior Rojas   u1114542   u1114542@utah.edu
      Saman Sepehri Nejad u1113773   s.sepehrinejad@utah.edu
  Github   **:** https://github.com/samansepehri/US-Mass-Shootings

- **Background and Motivation**

In last 50 years, the United States has had over 300 mass shootings that resulted in over 1900 deaths and over 2400 injured. The death tolls and places are different in every case, and the most recent mass shooting in Las Vegas had more than 500 victims, a very tragic incident with an uncommonly high number of victims. We believe that knowing more about the details of these incidents is of interest to many people like ourselves, who are currently living in the United States.

- **Project Objectives**

We would like to get some insights about common patterns in past US mass shootings. In particular, some interesting questions we would like to answer and some quantities and patterns we would like to show include:

  - Number of people killed and injured per year
  - Visualize the location of the incidents on a US map
  - Are these incidents more common in some states than others?
  - Are these incidents more common in certain months or days of the week?

- **Data**

Our main data source is www.kaggle.com, which provides a CSV file with information of US mass shootings since 1966.
URL: https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years/data

- **Data Processing**

The data requires some processing to be used in our visualization. There is no "state" column in the data set, but it can be extracted from the "location". There is no longitude and latitude in the dataset, but we can get the city where the mass-shooting happened, so that we can use that to locate positions on map. The "date" column is a simple string, which is not appropriate for certain comparisons and analyses we want to do. We will extract the year, month and date from this column. Since we also want to analyze patterns about days of the week, we will need to extend the data with the corresponding day of the week in which each incident happened. We also need certain aggregate quantities, such as number of victims per state and total number of victims per year,

which are not directly included in the data set. We will either pre-compute some of these values or compute them on the fly with JavaScript.

- **Visualization Design**

We guide our design by the general rule "overview first, zoom and filter, then details on demand". The following sections describe some proposals to achieve an effective visualization following this idea.

Overview by year and location

We propose that the initial state of the visualization show a bar chart with the total number of incidents per year. The years are selectable. When the user selects a year from the bar chart, the visualization is updated with a map that shows more information about the selected year. Allowing the user to select more than one year might also be a useful addition.

We propose two options for the map:

A. A tile map that shows each state and uses saturation as the channel to indicate the number of incidents.
B. A more traditional US map that shows the exact location of the incidents.
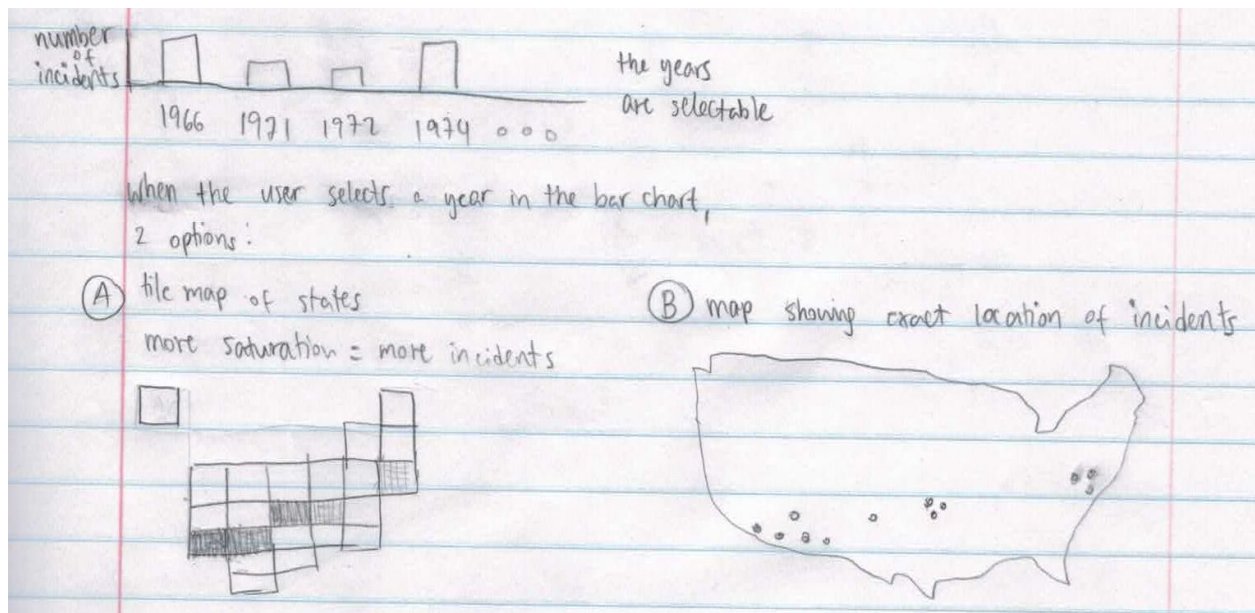
We believe both map options are valuable in different cases. Map A is useful for a quick overview of the number of incidents per state. Map B shows more detailed information of the locations and allows nice zoom and filtering, which is especially useful if most incidents happened in the same state. One important thing to note is that some incidents in our data set

do not have exact location information available, so for some years in which exact location information is scarce, Map A might be more appropriate.
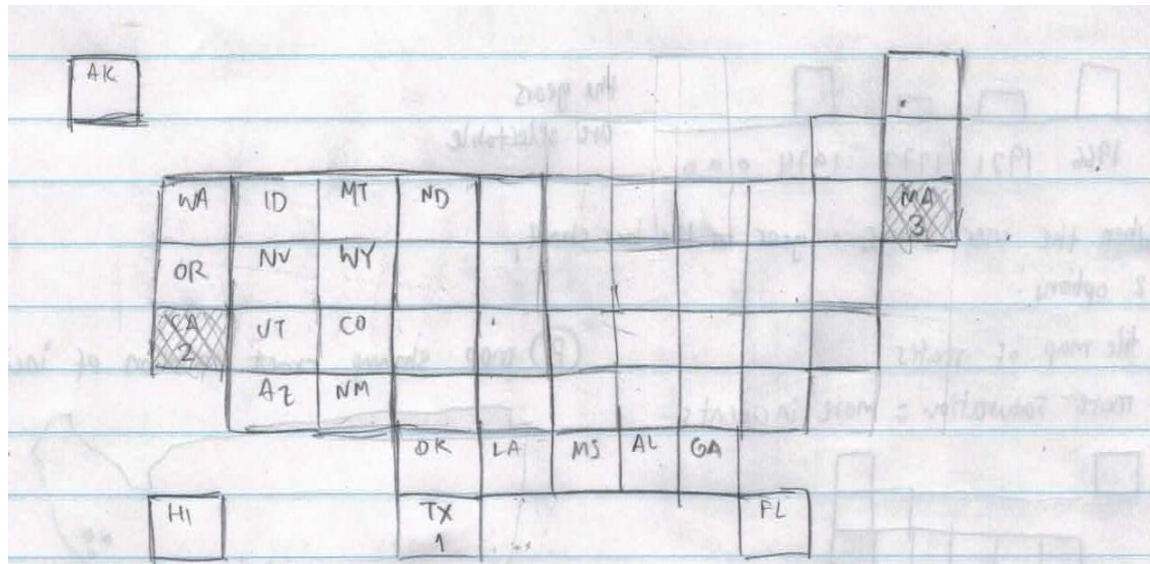
Another interesting alternative for an overview of the data per year is a bar chart of the number of victims, where each bar has two components: injured and killed, as shown in the following chart. In some cases, visualizing the total number of victims per year provides interesting insights that are absent in a chart of total number of incidents. For example, one single incident in 2017 had more than 500 victims, and the number of victims per year is often less than 50.
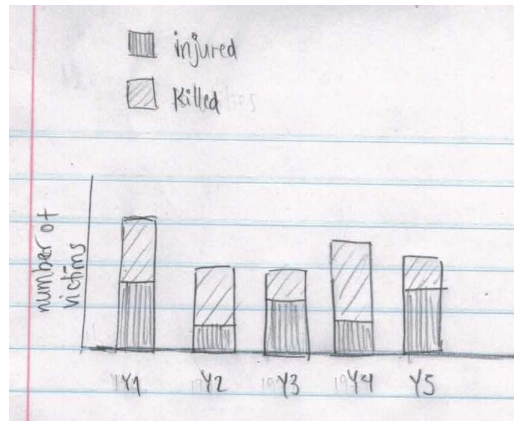
Filtering and details

Each map alternative proposed in the previous section provides different ways to filter the data and get more information about specific incidents.

Map A: The user can select states from the tile map, which in turn displays a table of detailed information of the incidents in the selected states.

AK											MA 3	
WA	ID	MT	ND									
OR	NV	WY										
CA 2	UT	CO										
	AZ	NM										
			OK	LA	MS	AL	GA					
HI			TX 1				FL					

Selected States: CA, MA

Title	Location	Date	Fatalities	Injured	...
title 1	San Bernardino, California	01/01/2016	1	5	
title 2	San Bernardino, California	02/02/2016	2	4	
title 3	Boston, Massachusetts	03/03/2016	3	3	...
title 4	Chelsea, Massachusetts	04/04/2016	4	2	
title 5	Boston, Massachusetts	05/05/2016	5	1	

Map B: This map uses circles as marks for each incident. The size of the circle can be used as a channel to encode a relevant criterion such as number of victims, and the position of the circle can be used as a channel to encode the city where the incident happened. When the user points to a circle, a tooltip with detailed information is shown.
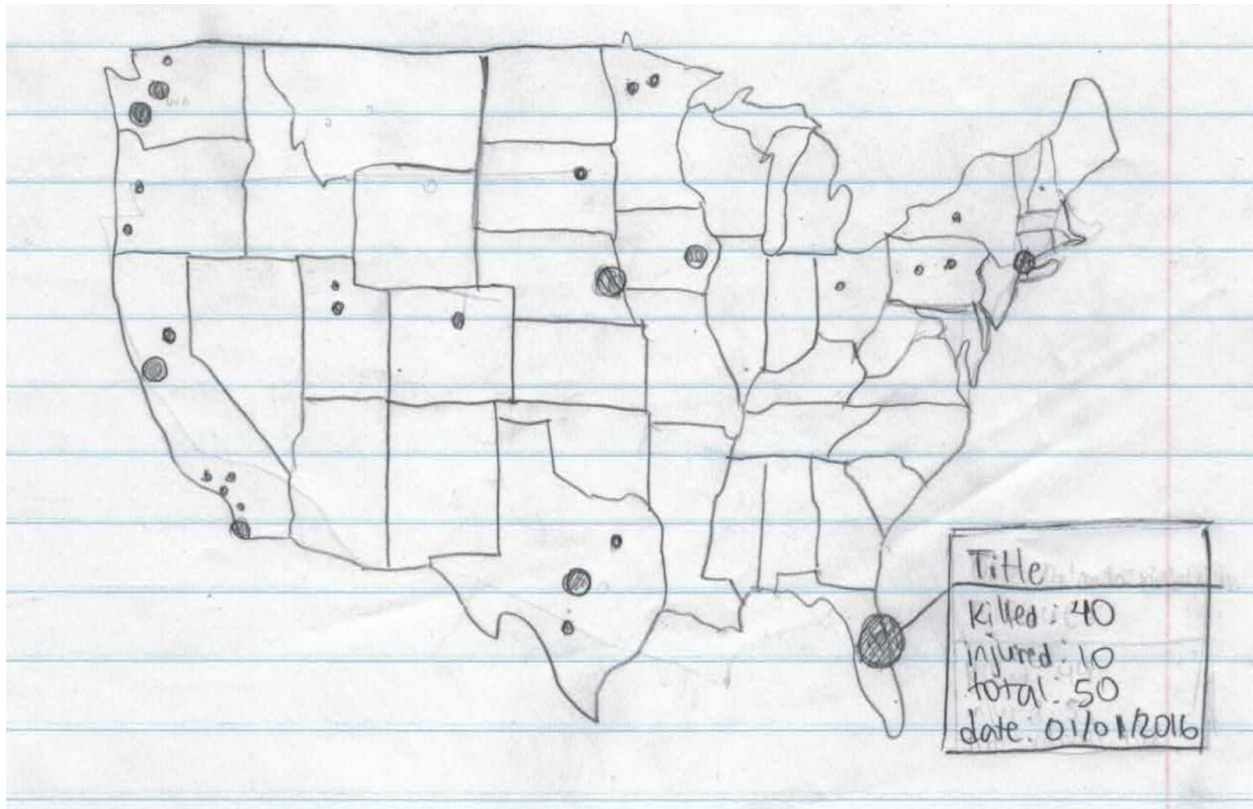

Comparison of changes over time in different states

We also want to get some insights about changes over time for different states. One alternative to achieve that is with a graph like the following, which shows the percentage of incidents in different years for different states. Other quantities of interest that could use the same type of graph are the number of people injured, killed and total victims. This type of graph is useful for our data because in some cases, plotting absolute values is not be the best choice. For example, one incident in 2017 had 585 victims, and the number of victims in one year is often less than 50. Comparing percentages is more appropriate for cases like this.

By number of incidents   Killed   injured

100%

0%

Y1   Y2   Y3   Y4   Y5

CA
MA
NY
IL
CA
MA
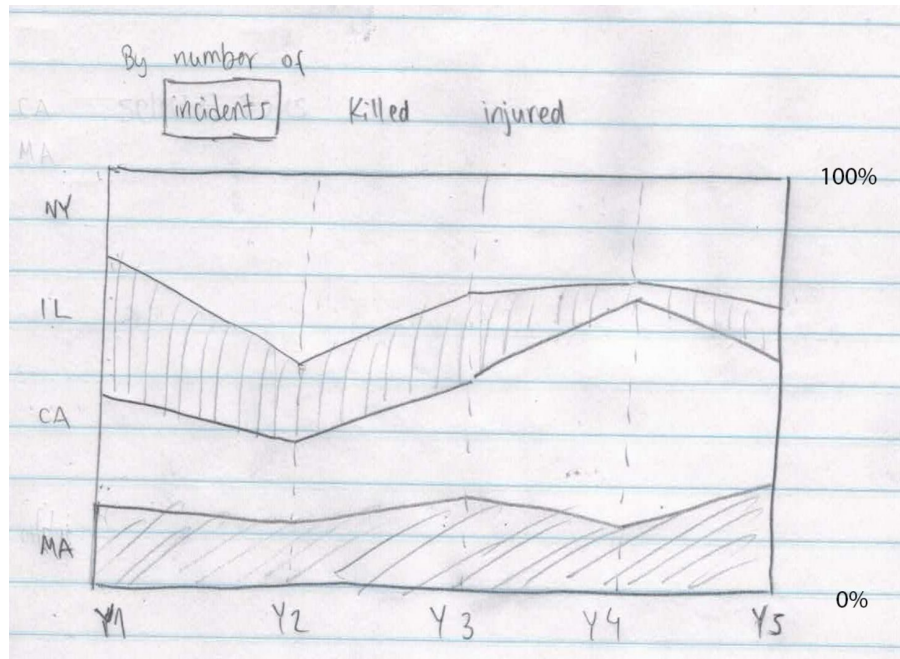
Scatter plot, killed - injured

It would also be interesting to get some insights about correlations between number of people injured and killed in each incident, for which we propose a scatter plot. As shown in the following plot, each point corresponds to an incident and the color encodes the total number of victims. A tooltip is also available when the user points to an incident.

50

injured 25

total number of victims

100

Title
Killed: 27
injured: 8
total: 35

Killed   25   50
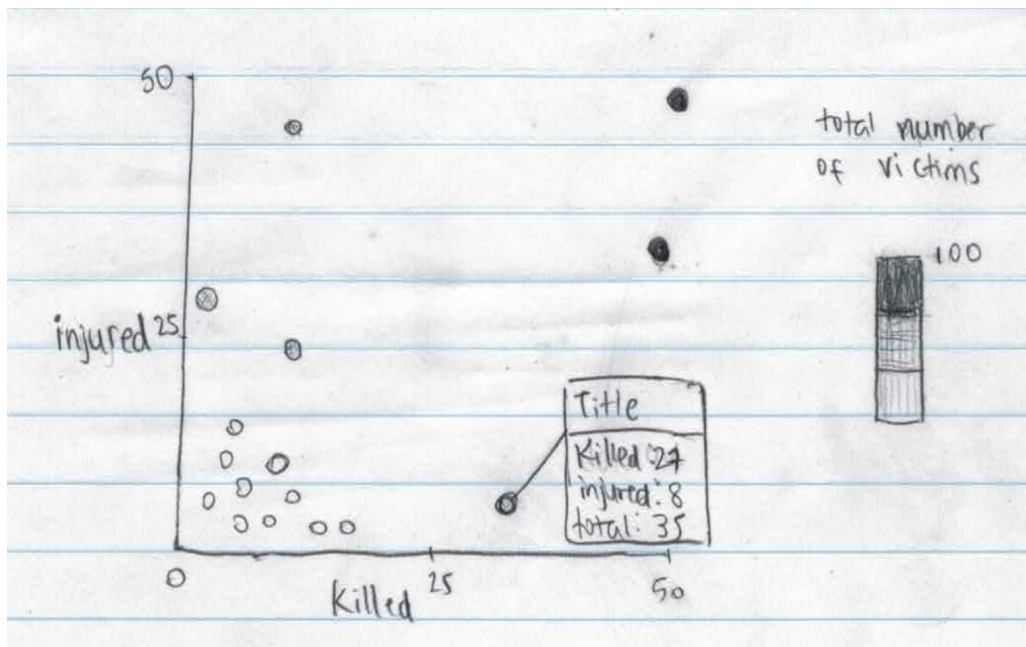
Patterns per month and day of the week

To analyze if some months or days of the week are more common for these incidents, we propose two alternatives shown in the following figures. One alternative is to show circles of different size (proportional to the number of incidents) for each combination of year - month and year - day of the week (the graph only shows months, but the same design applies to days of the week) in a grid-like arrangement. The other alternative is a bar chart by month (and day of the week), which provides a more concise representation, but the information per year is lost.
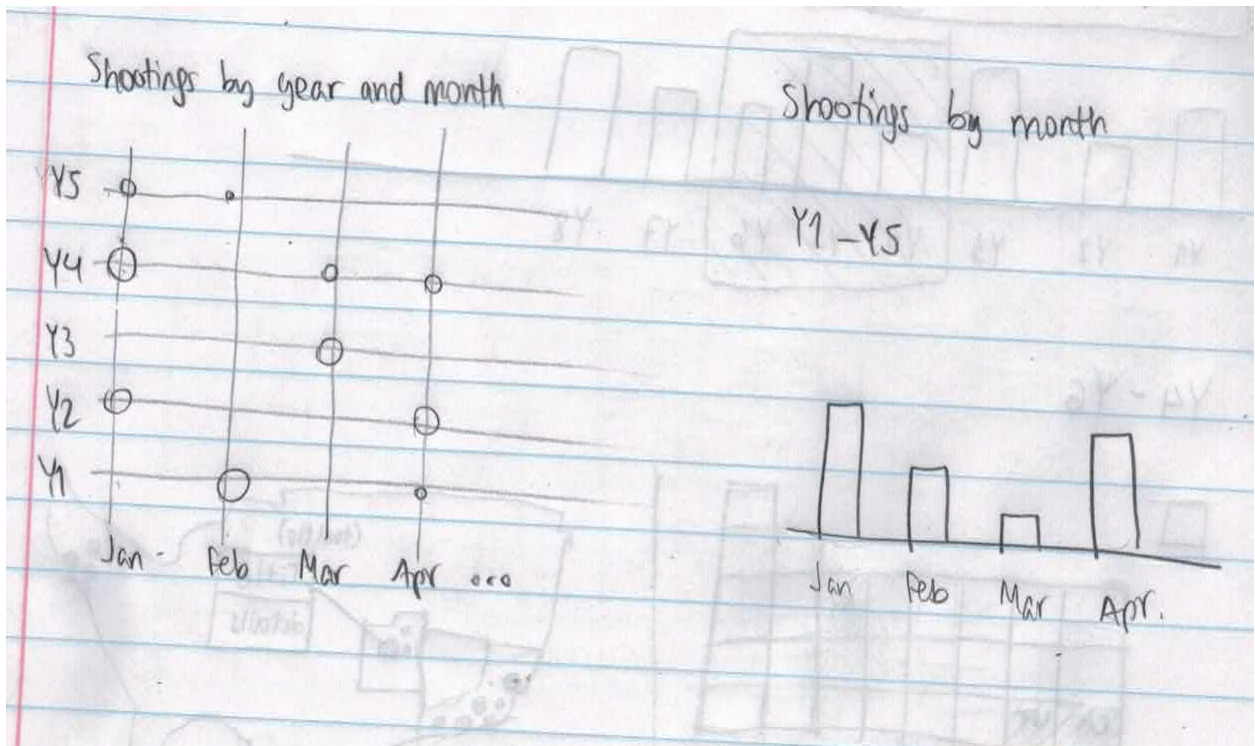
Proposed interactive visualization incorporating previous designs

The initial state of the visualization is an overview bar chart by years. The next level is to display a map for the selected years. The next level allows the user even more filtering, depending on the type of map (option A or B).

In addition, the visualization displays a comparison between different states and a scatter plot of injured - killed. The visualizations of frequency per month and day of the week are not shown in the following figure due to space limitations, but it should also be included in some part of the layout.

US mass shootings

Number of incidents ▼
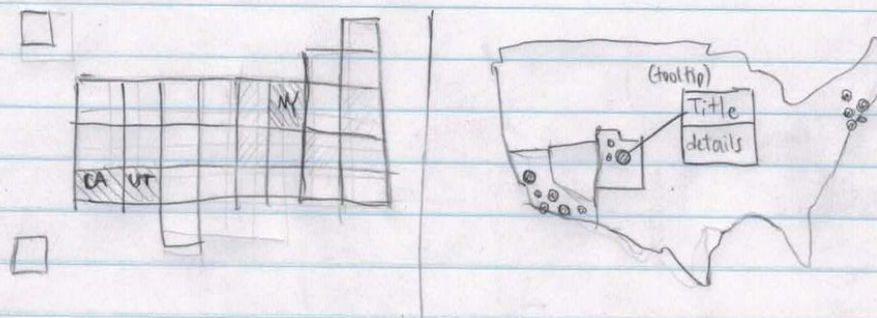

Y1 Y2 Y3 Y4 Y5 Y6 Y7 Y8

Y4 - Y6

NY

CA UT

(tool tip)
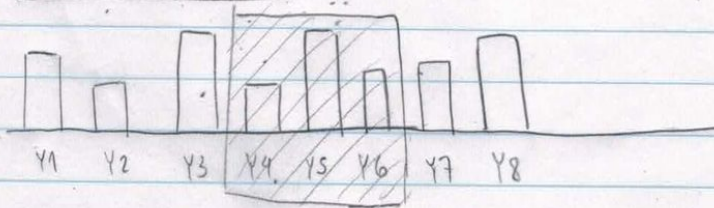
Title

details

Incidents (Y4-Y6)

Title	Location	Killed	injured	total	more details...
title 1	Salt Lake City, UT	1	4	5	
title 2	Provo, UT	2	3 ..	5	more details...
title 3	Los Angeles, CA	3	2	5	
title 4	San Francisco, CA	4	1	5	

% number of incidents ▼

Scatter plot

CA

NY

UT

Y4 Y5 Y6

50

injured

killed   50

total number
of victims

100
70
30
0
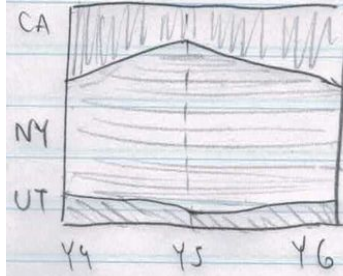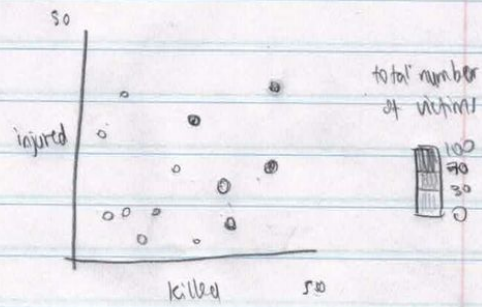
● **Must-have features**

  - Overview bar chart per year (either number of incidents or number of victims)
  - At least one map from the proposed options
  - The map must be updated to show information about the selected year
  - Table of incidents
  - At least one graph that compares some quantity (such as incidents or victims) between states.
  - Scatter plot killed - injured
  - Visualization of frequency by month and day of the week (at least one of the proposed options)

● **Optional features**

  - Allow the user switch between number of incidents and number of victims in the overview bar chart
  - Allow the user switch between different map options (tile map and geographical map) or show them side to side
  - Allow the user select more than one year
  - Allow the user switch between different comparisons (such as incidents or victims) between states
  - Allow the user rearrange the layout
  - Make the table of incidents sortable by column
  - Allow the user switch between different visualization alternatives for frequency by month and day of the week.

● **Project Schedule**

| Action Points	Oct	Nov				Dec
	Oct 28	Week 1	Week 2	Week 3	Week 4	Dec 1
Project Proposal	■					
Data processing		■				
Bar chart by year		■				
Table of incidents		■				
Map		■	■	■		
Project milestone			■			
Comparison between states				■	■	
Frequency by month and day				■		
Scatter plot				■		
Tweak interactivity				■	■	
Optional features					■	
Final submission						■

- **Data processing**

In order to get the correct location for each incident, we had to process the rows in the CSV file and add another attribute to specify the state names (abbreviations). There was some missing information for some entries. For example, some rows did not have (lat, long) information. Besides, the location values for different rows were not consistent. Some of them had only the name of the town or city as the location. We also had to remove 4 corrupted rows from the original data set because the number of fields did not match the number of fields of the data set and they missed some important information such as location, date and number of victims.

The format of the date field in our original CSV file is "mm/dd/yyyy". Since one of our charts is a bar chart of frequency by day of the week, we parsed the date of each row and computed the corresponding day of the week to extend the data set. Parsing the date was also necessary for our main bar chart by year. Finally, to get the list of states with their positions in our tile map (row and column) we used the 2012 election CSV files we had from HW6.

- **Peer review**

Feedback from other people helped us decide to choose between different options we designed in our original proposal. From the feedback we got from the peer review in class, and also from talking to the TA, we decided that the tile map would be a better option to visualize the location of the incidents (highlighted in yellow below), since it is more concise and showing the exact location of the incidents on a geographical map does not add much value.
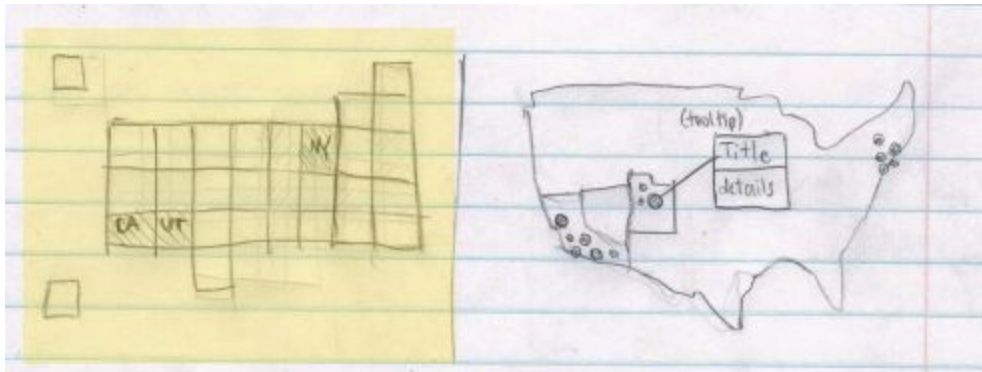
We also got feedback about having another visualization which can be useful to illustrate the relation between the population in an area and the number of incidents that happened on that specific area. The problem we found is that such information is missing in our original CSV file. We need to get the demographic data for each year in order to have this visualization. After discussing with the TA we decided not to include that in our project.

Thanks to the external feedback, we also realized that from the two options we proposed for visualizing trends by month / day, the bar chart (highlighted in yellow below) is more concise and useful, which is the option we decided to implement in our first prototype.
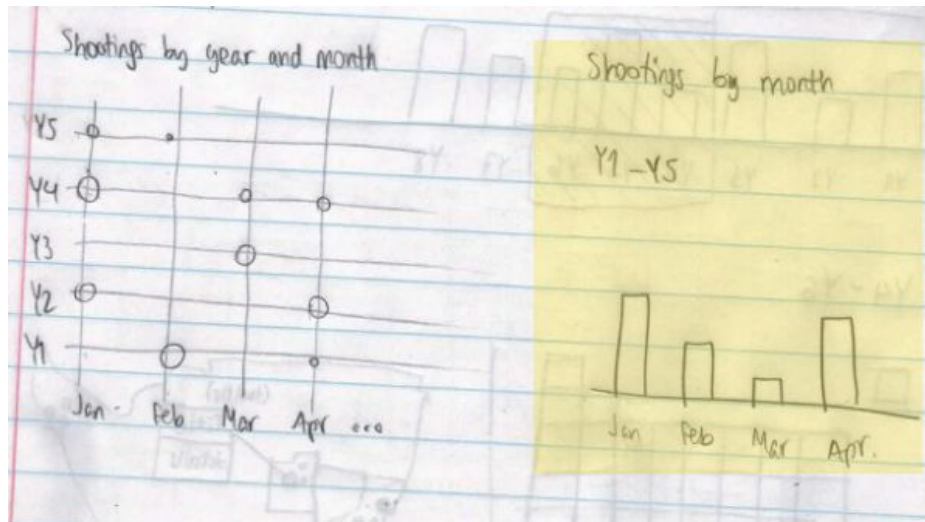
- **Prototype implementation for first milestone**

Bart chart by year
Following our proposal, the first visualization in our prototype is a bar chart of the number of incidents per year. We decided to rotate the year labels by 90 degrees to save space and be able to show all the years. A selection menu is available to switch between different criteria. For example, the figure below shows the number of incidents per year, but the interface allows switching to other criteria such as number of people injured, killed and total victims. In the final version, we will also have brushing feature to let users select a time frame and see the the rest of the charts being updated accordingly.
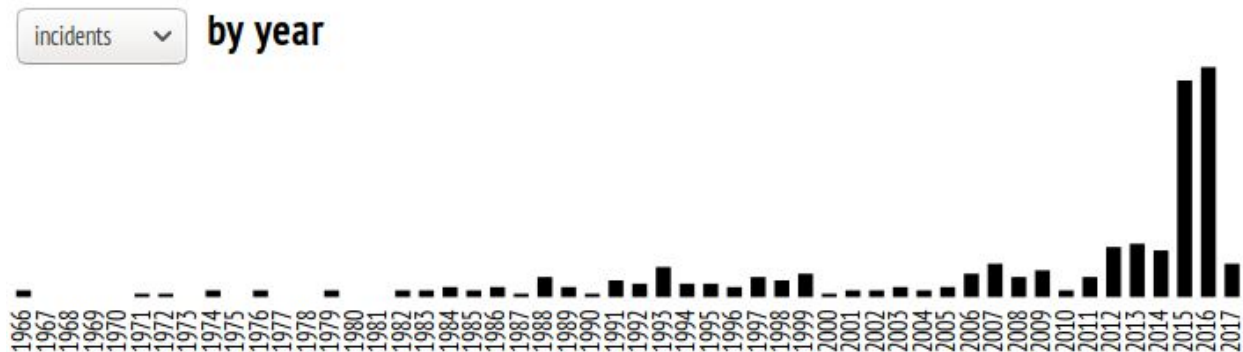
Tile map with information per state
Our prototype also includes a tile map that displays quantities (number of incidents, people injured, killed) per state, as shown below. For the chart below, color encodes the total number of victims (people injured + killed). We are going to have more features for the tile map such as hovering the

states to see the tooltip with detailed information and allowing users to select different states to narrow down the data for the other charts.
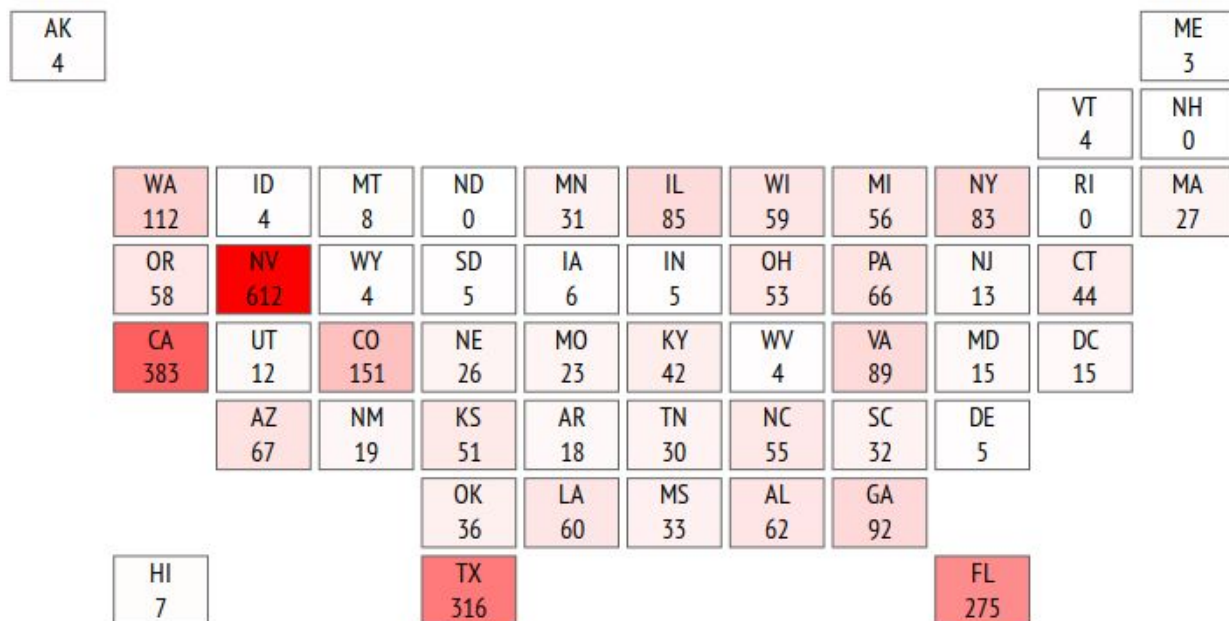
|										
AK 4									ME 3	
								VT 4	NH 0	
WA 112	ID 4	MT 8	ND 0	MN 31	IL 85	WI 59	MI 56	NY 83	RI 0	MA 27
OR 58	NV 612	WY 4	SD 5	IA 6	IN 5	OH 53	PA 66	NJ 13	CT 44	
CA 383	UT 12	CO 151	NE 26	MO 23	KY 42	WV 4	VA 89	MD 15	DC 15	
	AZ 67	NM 19	KS 51	AR 18	TN 30	NC 55	SC 32	DE 5		
			OK 36	LA 60	MS 33	AL 62	GA 92			
HI 7			TX 316					FL 275		

Table of incidents

Our first prototype also has an implementation of a table of incidents. In the final version, it would be linked with other views so that the list of incidents can be updated according to the selected years, states, etc.

Title	Date	State	Killed	Injured	Total Victims	Location	Area
Texas church mass shooting	2017/11/05	TX	26	20	46	Sutherland Springs, TX	Church
Walmart shooting in suburban Denver	2017/11/01	CO	3	0	3	Thornton, CO	Wal-Mart
Edgewood businees park shooting	2017/10/18	MD	3	3	6	Edgewood, MD	Remodeling Store
Las Vegas Strip mass shooting	2017/10/01	NV	59	527	586	Las Vegas, NV	Las Vegas Strip Concert outside Mandala Bay
San Francisco UPS shooting	2017/06/14	CA	3	2	5	San Francisco, CA	UPS facility
Pennsylvania supermarket shooting	2017/06/07	PA	3	0	3	Tunkhannock, PA	Weis grocery
Florida awning manufacturer shooting	2017/06/05	FL	5	0	5	Orlando, Florida	manufacturer Fiamma Inc.
Rural Ohio nursing home shooting	2017/05/12	OH	3	0	3	Kirkersville, Ohio	a nursing home
Fresno downtown shooting	2017/04/18	CA	3	0	3	Fresno, California	a street in downtown
Fort Lauderdale airport shooting	2017/01/06	FL	5	6	11	Fort Lauderdale, Florida	baggage claim area of the airport

Incidents by day

We also implemented the functionality required to compute and display quantities (such as number of incidents) by day of the week, as shown below.

Incidents by day

| mon | tue | wed | thu | fri | sat | sun |
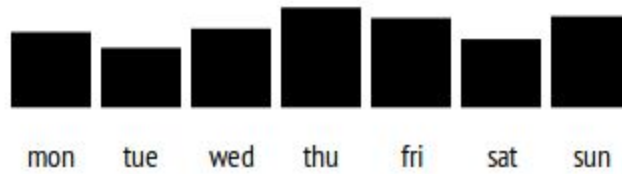
Scatter plot killed-injured

As designed in the original proposal, we also implemented the scatter plot killed-injured for our first prototype, as shown below. Each circle represents an incident and the saturation is proportional to the number of total victims (killed + injured).

Scatter plot, killed-injured
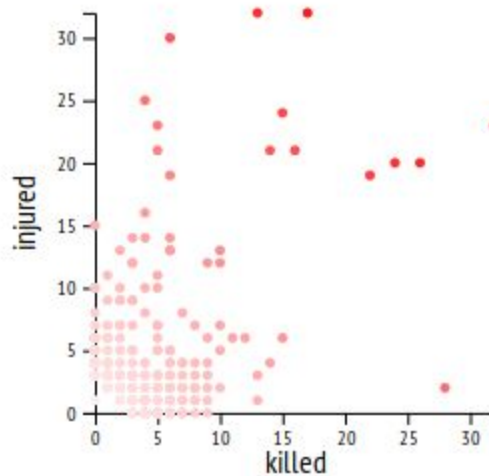
Comparison of states through time

The following is a sample of the implemented prototype using fictitious data for the graph we designed for the initial proposal to compare the percentage of incidents in different states for different years. For each year, each state has a fraction of the vertical space that represents the percentage of incidents associated to that state from the total number of incidents during that year in the selected states. The example below compares three states over four years.

States through time
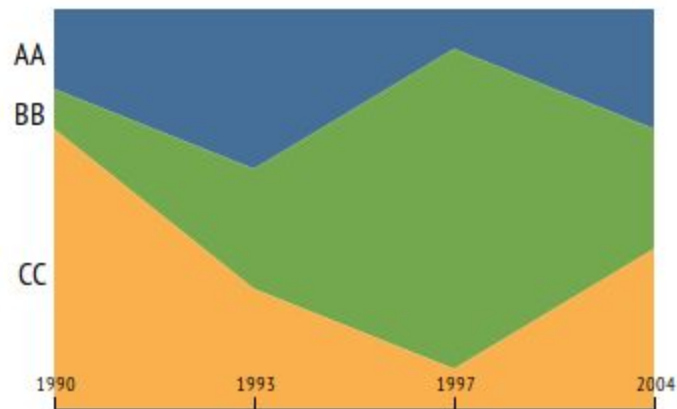
Although this seemed like a reasonable representation at first, we found some issues when using numbers from our real data set. It is not uncommon for a state to have zero incidents in many years, which causes undesired overlaps and breaks the effect of continuity, as seen in the example below.

States through time
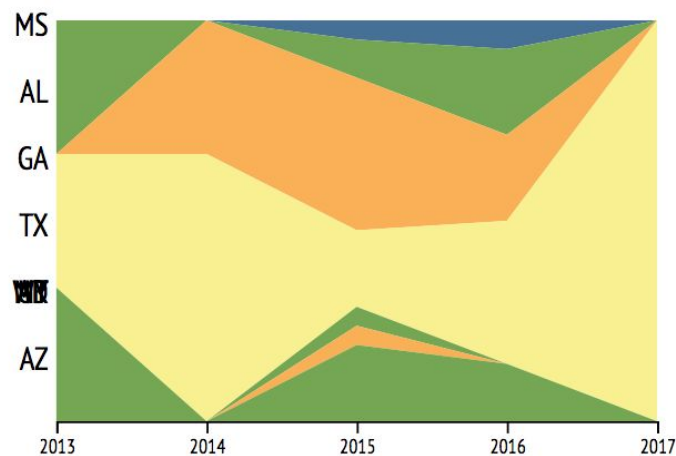
We will consider other designs, such as line charts, which might produce a better visual result.

Overall view of the implementation (first milestone)

US mass shootings

incidents **by year**

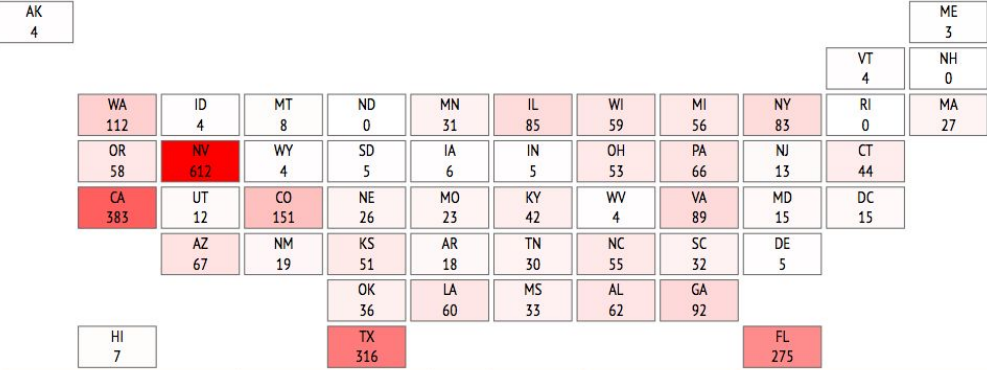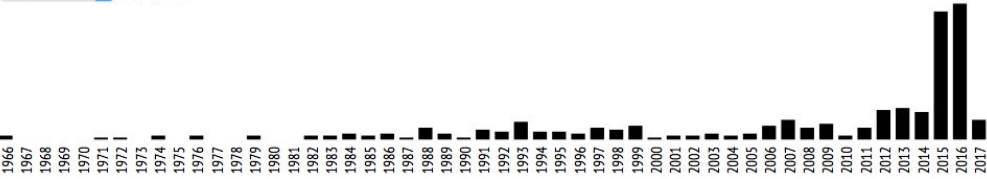|										
AK 4										ME 3
									VT 4	NH 0
WA 112	ID 4	MT 8	ND 0	MN 31	IL 85	WI 59	MI 56	NY 83	RI 0	MA 27
OR 58	NV 612	WY 4	SD 5	IA 6	IN 5	OH 53	PA 66	NJ 13	CT 44	
CA 383	UT 12	CO 151	NE 26	MO 23	KY 42	WV 4	VA 89	MD 15	DC 15	
	AZ 67	NM 19	KS 51	AR 18	TN 30	NC 55	SC 32	DE 5		
			OK 36	LA 60	MS 33	AL 62	GA 92			
HI 7			TX 316				FL 275			

Title	Date	State	Killed	Injured	Total Victims	Location	Area
Texas church mass shooting	2017/11/05	TX	26	20	46	Sutherland Springs, TX	Church
Walmart shooting in suburban Denver	2017/11/01	CO	3	0	3	Thornton, CO	Wal-Mart
Edgewood businees park shooting	2017/10/18	MD	3	3	6	Edgewood, MD	Remodeling Store
Las Vegas Strip mass shooting	2017/10/01	NV	59	527	586	Las Vegas, NV	Las Vegas Strip Concert outside Mandala Bay
San Francisco UPS shooting	2017/06/14	CA	3	2	5	San Francisco, CA	UPS facility
Pennsylvania supermarket shooting	2017/06/07	PA	3	0	3	Tunkhannock, PA	Weis grocery
Florida awning manufacturer shooting	2017/06/05	FL	5	0	5	Orlando, Florida	manufacturer Fiamma Inc.
Rural Ohio nursing home shooting	2017/05/12	OH	3	0	3	Kirkersville, Ohio	a nursing home
Fresno downtown shooting	2017/04/18	CA	3	0	3	Fresno, California	a street in downtown
Fort Lauderdale airport shooting	2017/01/06	FL	5	6	11	Fort Lauderdale, Florida	baggage claim area of the airport

Incidents by day

States through time

Scatter plot, killed-injured

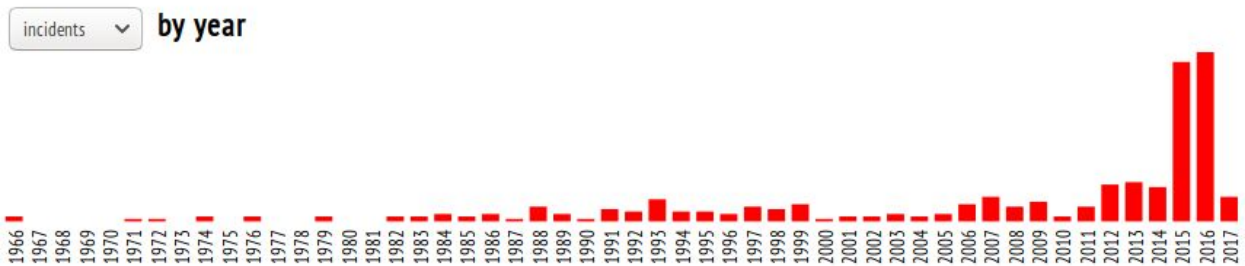**Process after first milestone**

Year chart interactivity

The year chart is the main entry point of our visualization. It shows the total number of incidents per year and our original design proposed to allow the user to select one year to update the other charts using only data from the selected year.

Bar chart of incidents for all years

While interacting with the visualization, we realized that adding support for year range selection in addition to single years was very valuable. As shown in the last figure, it is difficult to compare the data before 2015, because of the scaling caused by the number of incidents in 2015 and 2016. To fix this issue, we implemented a feature that allows the user to select a year range by dragging over the year labels. Every time the year range is updated, the scale of the chart is updated as well, as shown in the following figure.

Bar chart of incidents with updated scale for selected year range
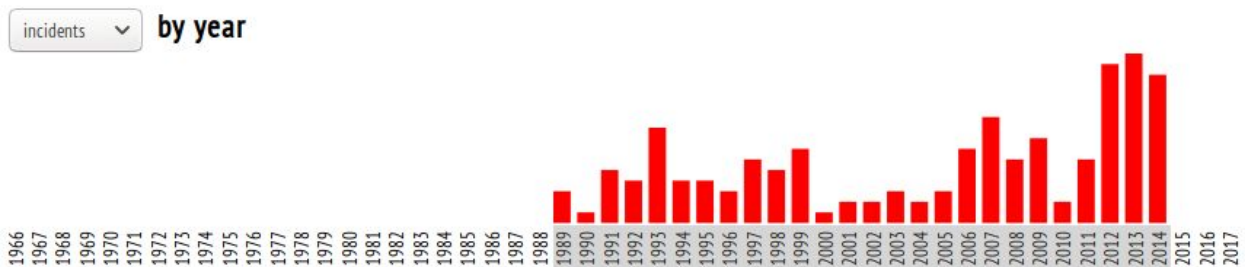
The criterion of the bar chart can also be updated. By default, it shows the number of incidents, but it can be switched to number of injured, killed and total victims. In particular, it is interesting to switch between number of incidents and total victims and realize that sometimes the years with more incidents are not always the years with more victims. For example, as shown in the following figures, between 2012 and 2017, although the year 2017 has the least amount of incidents, it has the highest number of victims.

Incidents between 2012 and 2017. The year 2017 has the least amount of incidents.

**Number of victims between 2017. The year 2017 has the highest number of victims.**

Details on demand

Once we identify an interesting pattern like the one we showed in the previous figures with data between the years 2012 and 2017, we can analyze the other charts since everything is linked. For example, thanks to the other charts in our visualization, we can easily identify that the high number of victims in 2017 is due to a single incident in Las Vegas, as shown in the following figure.

Number of victims between 2012 and 2017. The number of victims in 2017 is due to a single incident in Las Vegas.
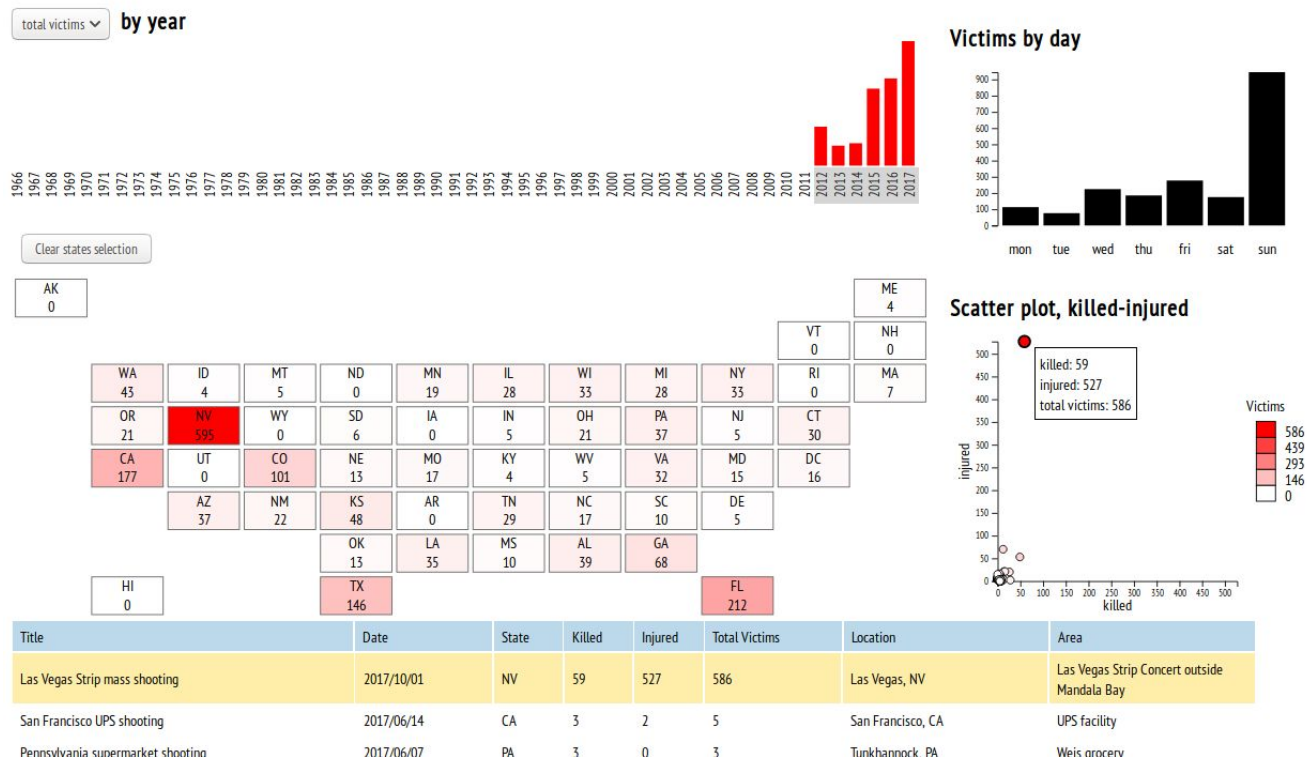

In particular, the scatter plot is a useful tool to identify rare events like the incident in Las Vegas, as shown, in the last figure. While the year chart, day chart and tile map are good tools to find general patterns in time and different states, the scatter plot and the table of incidents are good tools to explore individual events. When the user points to one incident in the scatter plot, it gets highlighted in the table, and vice-versa, as shown in the following figure.

The scatter plot and table of incidents are linked. When an incident is selected in the scatter plot, it gets highlighted in the table and vice-versa.

Scatter plot, killed-injured
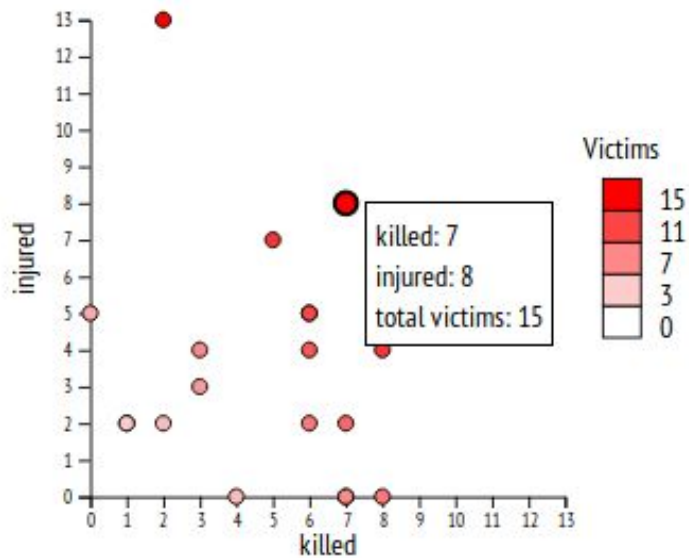
Title	Date	State	Killed	Injured	Total Victims	Location	Area
Birchwood, Wisconsin	2004/11/21	WI	6	2	8	Birchwood, Wisconsin	forests of Birchwood
Windy City Core Supply Warehouse	2003/08/27	IL	7	0	7	Chicago, Illinois	Warehouse
Lockheed Martin shooting	2003/07/08	MS	7	8	15	Meridian, Mississippi	Workplace
Case Western Reserve University	2003/05/09	OH	1	2	3	Cleveland, Ohio	School
University of Arizona College of Nursing	2002/10/28	AZ	4	0	4	Tucson, Arizona	University of Arizona
Appalachian School of Law	2002/01/16	VA	3	3	6	Grundy, Virginia	Appalachian School of Law
Santana High School	2001/03/05	CA	2	13	15	Santee, San Diego, California	Santana High School
Navistar International	2001/02/05	IL	6	4	10	Melrose Park, Illinois	Navistar in Melrose Park
Wakefield massacre	2000/12/26	MA	7	0	7	Wakefield, Massachusetts	Edgewater Technology

Comparison of states over time

After implementing the prototype for the first milestone, one of the main things we realized was that our proposed design to compare the number of incidents in different states over time was not appropriate for our data. As shown in the following figure, our original idea was to display one area per state by connecting segments of size proportional to the number of incidents per year. The issue with this design was that states with no incidents in several years was not an uncommon pattern, which caused some shapes to shrink to zero height.
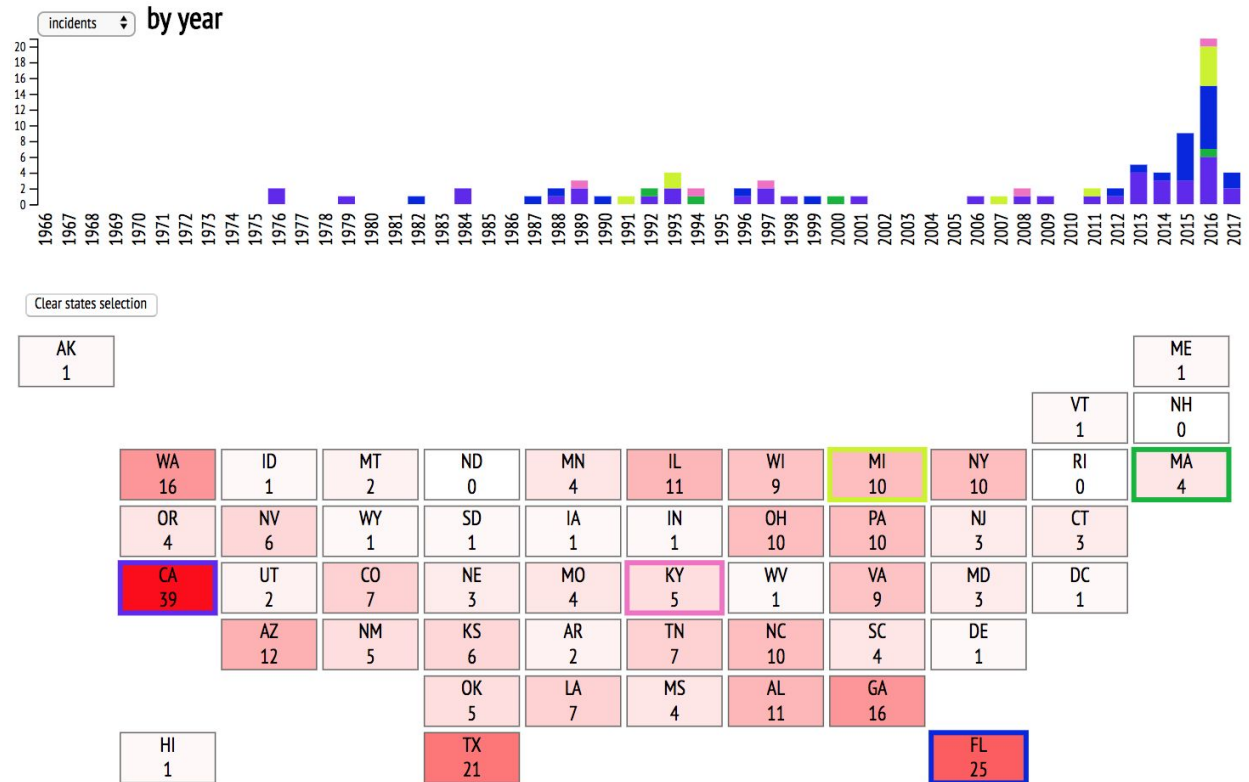
Initial prototype to compare states
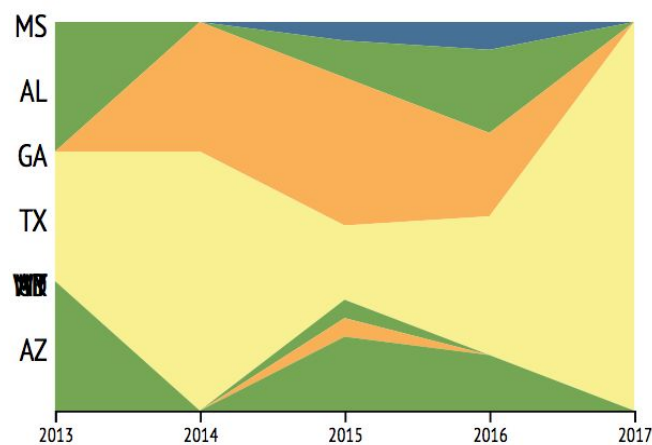
To fix this, we decided to re-use our year chart, by coloring different portions of each bar. The user can select states from the tile map, which updates the year chart as shown in the following figures.

Updated year chart showing data from different states for all the years

incidents ⇅ **by year**

Clear states selection

AK 1											ME 1
										VT 1	NH 0
WA 16	ID 1	MT 2	ND 0	MN 4	IL 11	WI 9	MI 10	NY 10	RI 0	MA 4
OR 4	NV 6	WY 1	SD 1	IA 1	IN 1	OH 10	PA 10	NJ 3	CT 3	
CA 39	UT 2	CO 7	NE 3	MO 4	KY 5	WV 1	VA 9	MD 3	DC 1	
	AZ 12	NM 5	KS 6	AR 2	TN 7	NC 10	SC 4	DE 1		
			OK 5	LA 7	MS 4	AL 11	GA 16			
HI 1			TX 21				FL 25			

Updated year chart showing data from different states for a range of years

incidents ⇕ **by year**