US Mass Shootout Visualization

SUPREET SINGH VIRDI

STUDENT NUMBER 17312704

Abstract

This paper present the approach used in the visualization of the dataset, and denotes the various encoding channels used to make the interactive view. Further providing the Choropleth approach to get the thematic map in which area are shaded to conclude the statistics of the dataset. Interactive approach to the set of data debriefs the impacted location on the map providing the detailed information. This overall concludes in the infographic for the statistical variable in the dataset describing the increase in the rate of incidents over the years and highlighting the sensitive shootouts.

1. Introduction

This visualization is created in the D3.Js using the United States mass shootout dataset which is downloaded from the Kaggle web-site. This visualization is created for the coursework of CS7DS4 module, which depicts the representation of visualization tech-niques covered in the course structure.

United States Mass Shootout dataset lies under the category of the Crime dataset. This specifically focuses on the shootout inci-dents occurred in the different places of the US over the past years. This dataset contains more than 300 incidents occurred in USA and contains more than 20 features from year 1966 to 2016. It tells the incident details such as summary of the incident, cause, location, date, victims etc. There are few missing data in the dataset so lit-tle bit pre-processing of the data is required so that the information from the data cannot be tampered and a meaningful visualization can be created. This dataset gives the view to the readers about the clear area affected with the specified date and location and the most important it gives victims affected. This overall gives the interactive visualization, which defines the choropleth map by using the differ-ences in the shading and color within pre-defined area to indicate the highest rate in the states of the United states of America.

2. Approach

This visualization is created using choropleth with the interactive approach by plotting the data point based on the years and various others factors. The map is divided among depicting the different shading of the color on a scale of 0 to 30 to represent the number of the shoot out incidents occurred in every state. The darker colored state shows the highest number of shootout in the states. To repre-sents this, various spatial marks and encoding channels are used to specify the location such as pinpoint mark and differentiating the regions such as color-coding.

To create the map USA JSON file is used, which contains the state wise coordinates and the overall coordinates of the USA. Then using the TopoJSON, I have encoded the topology rather than rep-resenting in the geometrics, these are stitched together by an arc. SVG function is created to load the coordinated from the JSON file and create the map. This file contains the multiple features collec-tions without any duplication such as states and counties. In parallel to this CSS is used to provide the spatial marks and the attractive color to the maps.

This visualization briefs the two information from the dataset with the conclusion. In the first it shows the most affected shoot out states where the number of shoot out is more. The dark color with #d65c5c shows the highest number of the shootout incident in the states, whereas it decreases to the light color to #f6dede with decrease in the number. The data is spread from year 1966 to 2016 covering 302 incidents. The JSON file consists of the unique id, which is particularly assigned to each states, using the unique id specific color has been assigned to each states that can be seen in the scale in the visualization.

Effectiveness and Expressiveness can be used in the visualiza-tion to express the set of facts of the information. Color encoding is expressing the information about the states proving the informa-tion about the incidents in the states. Along with this to provide the visual encoding channel such as spatial marks pinpoint is used to point the location of the incident using the latitude and longitude feature of the dataset. The size of the marks relates to the features i.e. total victims in the dataset, which provide the clear understand-ing to the reader about the impact of of the incident on the particular location. The larger size of the mark denotes the highest number of victim in the incident and the size decreases with the number of victims. Math function is used to define the size the of the marks by finding the square root of the each value of the total victim and then multiplying it with 10 to give the uniformity in the size of the

marks and to provide clear understanding. To give the clearer persona to the visualization, Visual Encoding channel such as motion used to implement the zoom feature and to get detailed information about each mark based on the region and pointed location.

Detailed representation is provided to the visualization by dividing the data into the tabular form, for the better representation and to get the non-populated view of the representation. The data has been divided into the range to satisfy that the visualization would not be densely populated. The data is processed into the table on the basis of batch by comparing the date form the dataset and further analyzing to fit the particular data in the respected batch to process in the tabular form with the specified id. Total 7 batches is created and compared by years feature in the loop. The range of data from year 1966 to 2016 is divided into a table of 6 rows i.e. 2015 to 2016, 2013-2014, 2009-2012, 2005-2008, before 2005 and from 1966-2016, which separates the data into various chunks. To give the meaning to the represented data, tooltip is used on mouse over to give the detailed knowledge of the dataset to the reader by showing the name of the shootout, location, date and the total victims. These all provides the interactive approach to the visualization, which can be seen in the Figure 1.

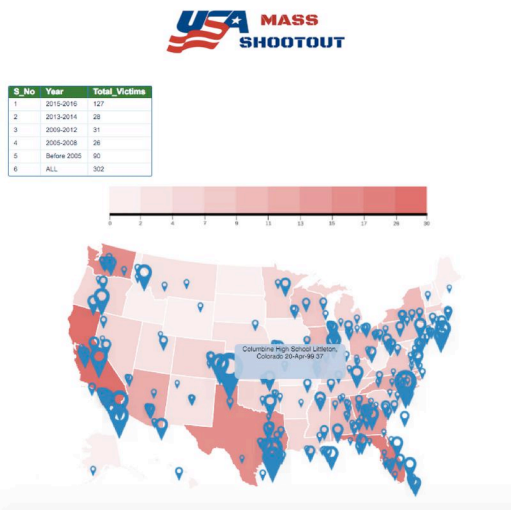

Figure 1: Visualization of the dataset.

The Infographic is also created to highlight the important information from the dataset, which provides the major incidents, happened in the past with the specific data. This provides the clear conclusion to the reader about the hypothesis showing the percentage of increased crime rate in the past years and pointing out the largest shootout in the United States of America, this highlights the important feature of the dataset such as the total incident occurred over the past years. This is created in the Canva Tool to provide the clear understanding and point the statistics of the dataset.The high-lighted top shootout represents the clear story of the incident which tells the exact date, location and total victims impacted because of the shootout, which can be seen in Figure 2.

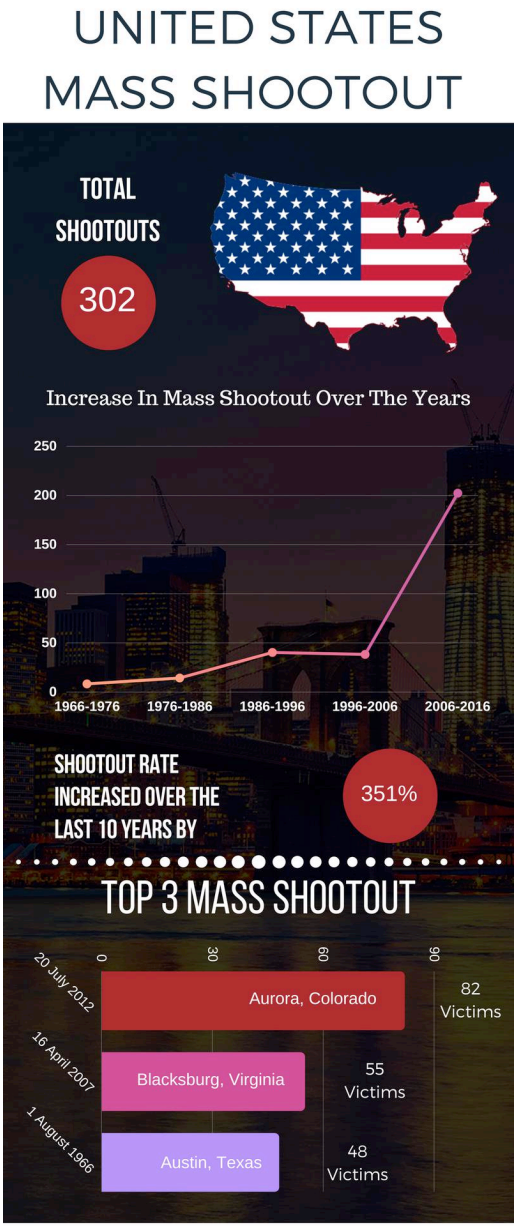

Figure 2: Infographic.

3. Conclusion

This visualization clearly concludes the increase in the crime rate or specifically shootout in the different region of the United States. From this visualization it can be concluded that the shootout incidents is increases by 351% in the last 10 years i.e. after 2006. And this shows that the major incident occurred with the highest number of victims in 3 three states i.e. Colorado, Texas, Virginia. Whereas by seeing in the Choropleth map it can be concluded that Florida, California, Texas and Washington are the states with the highest number of shootouts.

References

[1] https://bl.ocks.org/wboykinm/dbbe50d1023f90d4e241712395c27fb3
[2] https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years/data
[3] https://bl.ocks.org/mbostock/2374239
[4] http://bl.ocks.org/andreaskoller/7695784