Process book
CS 360: Data Visualization
Spring 2018
Sahil Jain - email
Rawan Al Mubarak - email
Project Repository: Repository
Project Website: Website

Project Title : Gun Deaths

This document is about the process that I followed while working over this project. The main purpose of this document is to keep track of all the processes we followed during the completion of this project as well as tasks that were accomplished. Process book will divided into five subcategories, first one will be motivation and goal behind this project, Second will be about data, third will be about visualizations, fourth will be about the challenges and fifth will be about analysis, challenges and tasks accomplished.

OVERVIEW AND MOTIVATION

Being an international student and living in United States we always found it very intriguing to work on the data related to the US. First time we got this opportunity in our liner regression class in Fall 2017 semester, but it was about doing regression analysis than visualizing and telling story through data. This semester we got a chance to work on a different data set that was related to gun violence in the US. As an immigrant living in US, we took this opportunity to make decisions and base our opinion on the bases of data not on the bases on political correctness.

Motivation for this project is based on the current events that have been taking place in the US over past few years related to gun violence and gun crimes the movement for strong gun laws and repealing 2nd amendment have been on the rise. The recent outburst has been caused by Florida high school shooting which leads to the death of several young individuals, and last year Las Vegas shooting, Texas church shooting was also some very disturbing events that lead to the need of stronger and stricter background check and gun control laws. Recently in California, in YouTube headquarters, a shooting incident took place as well, resulting in the injury to 3 people. Gun violence-related crime has been on the rise from past few months now and by visualizing data it can be clearly seen that whether gun violence crime over the span of years have gone down or have they rose in numbers. The motivation behind choosing this particular project is to provide solid evidence in favor of an argument. According to the 2nd Amendment “A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed” is a constitutional right for every American Citizen just like other laws. Clearly, the main aim of our project is to provide our own opinion through data but not to advocate against or in favor of gun laws, all our project will be trying to do it to tell a story about gun violence with help of visualizations and data.

RELATED WORK

There were plenty of online resources available who worked over this project, but the problem was most of these sources were very political and tried to made stories that would favor them rather than telling a real story behind the data. However I was able to get my way across plenty of nice and informative resources for my project. Below area the links posted to similar projects to our project.

(1) An article related to gun violence in Mother Jones magazine

Mother Jones I

Mother Jones II

(2) Article related to gun violence and homicide in The Guardian

The Guardian

(3) Gun Violence related visualization on Vox

Vox

(4) Gun Violence with relation to gun used, sex and race visualization

Gun Crime

(5) Analysis on gun death by CDC and Every town research

CDC Research

Everytown Research

QUESTIONS

Since the project that we were working upon was a very politically biased and sensitive topic asking a appropriate question was really important here. Not only one but many questions were supposed to be answered through these visualizations. Most important aspect of our project was the data correctness not political correctness, meaning answering questions based on the data not based on personal political preference. In beginning of the project our main question was "Should 2nd Amendment be repealed" but as the project continued forward we saw lots of different changes in our question from "fixing 2nd Amendment" our new question was "Is there

really a need for stricter gun laws?" and second question was "Is the US really that bad as media portrays it in terms of gun violence?". These questions evolved during the process of making visualizations and during general analysis of the data, while comparing the US with rest of the World gun violence data we noticed that US isn't that bad as media portrays, and while comparing different US states with each other form gun violence we came across that states with high gun control have higher shootings. So at the end instead of simply focusing on one question, we switched are focus to answer two or three multiple questions at the same time and those questions were "Is there really a need for stricter gun laws?" and "Is the US really that bad as media portrays it in terms of gun violence?"

DATA

Gathering data for this project was relatively easy however cleaning up data was immensely hard and took a lot more time as it was supposed to take. Data was downloaded from 4 sources in total, these sources were: (1) FBI, (2) Kaggle, (3) The Guardian (4) Mother Jones. FBI was used to download data about shootings in US from year 2003-2016, Kaggle, The Guardian was used to download mass shootings data in US lastly Mother Jones was used to download data related to gun violence across the world. (links posted in the related work section)

Data cleanup and look up was extensively done in R and Excel. Cleanup of the data consisted of processes like removing extra variables and NA values, adding extra variables like latitude and longitude value to show the data on the map and lastly joining 5-6 datasets together using SQL join command. US states data was crunched into one single file instead of 14 different files by number of shootings took place in each state in each year. In world gun violence data some of the countries were removed from the data for eg. China and Russia, because no data was available for those two countries. For mass shootings data, data was divided into 6 different data sets, division was initiated on the bases of places (school, workplace etc etc) where mass shootings took place and latitude and longitude values were added manually after an extensive search on google search engine. Lastly, for mass shootings data was further divided into even smaller data sets for the guns used during mass shootings and number of fatalities in each and every mass shooting since year 1982-2018.

Here's something interesting we came across while visualizing the data which we define as "data baiseness". This point is interesting in itself because when we were working on this project and while visualizing the data we noticed an unusual trend within US states shootings data. FBI describes shootings in assault and shootings in defense as two different category, however some

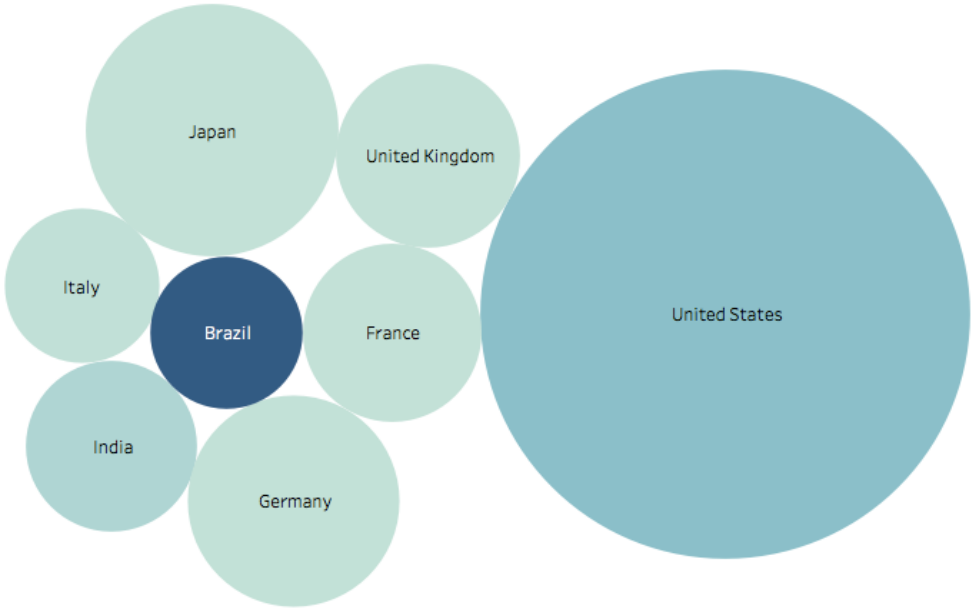
of these sources were using defense and assault as the same category, meaning tweaking the data little bit to make results go in there own favor.

EXPLORATORY DATA ANALYSIS

For Exploratory data analysis we already had a solid design in our minds but it took quite a while to came across the final design. In beginning we used tableau and R to make simple basic designs and made a whole mind map on how to go around with this whole project. Here's very first graph we made on tableau:

This world graph we made in tableau kicked start our project and helped us founding are way through this whole project. Here each and every country has Rank and Number of homicides by firearms respectively. Up next we kept on uploading the data in tableau and coming up with the final design of the project and made a good solid mind map on how to implement the design for this project. Here's another chart we made in tableau to get an insight to the data itself, where

bubbles are sized in term of the GDP of the country and color is on the basis of number of gun



homicides in each of these country.

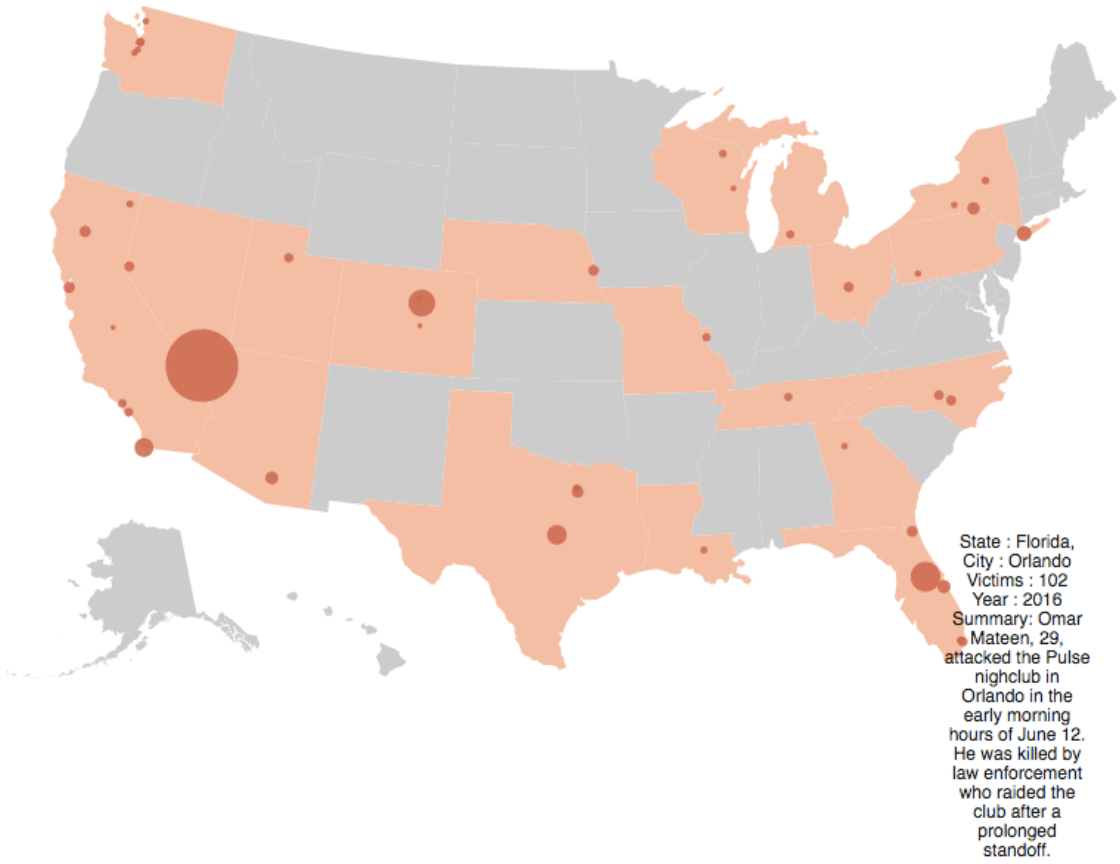
Our initial visualizations were simple bar graph, scatter plots and choropleth maps which evolved over time as the project kept moving forward. During these initial design is when we get know more about our data and that changes we are supposed to make in our data, also it gave us a long shot about how our final visualization design is going to look like. After doing some initial design analysis in the beginning we decided to stick to a strict design pattern without changing whole design itself.

DESIGN EVOLUTION

As our project kept moving forward, we noticed few changes that were taking place in our project. Over time our main question changed from simply addressing one question to addressing two questions by the end of the project as we gained more insight to our data. Change in the

approach to the project also changes our design for this project it evolved from more sort of a research to an overall analysis of this whole topic. Different visualizations we considered were multiple number of line charts, bubble charts, maps and bar charts but as the project continued we switched our focus from multiple number of visualizations to simple focus on more detail oriented visualizations. For example, we decided to use bubble chart or stacked bar chart while comparing top countries on the bases of Gun ownership and GDP, however while visualizing this specific soon we realized that gun ownership scale is way bigger than that of gun homicide, so we completely dumped the idea of using a stacked bar chart. Up next bubble chart, using bubble chart seemed like a very solid visualization design but developing X and Y axis in case bubble chart were equally hard, it was not like it's not possible, that part was certainly possible however even after multiple tries visualizations just wont look exactly the way we wanted them to look like. After rejecting both stacked bar charts and bubble charts, we decided to stick with more simpler trivariate scatter plots, bar charts, maps and line charts with interactions and details on demands.

We made this decision simply based on one fact and that was delivering clear message across the board. We could have made other fancier visualizations but they wont exactly portray the story that we were looking to tell using the data. Other thing that played part was user interaction in which a user can hover over pretty much any graph and will be provided with every necessary information needed related to the data and visualizations. Before beginning our project we were not really in favor of using "simple text" in our visualizations but over time it became clear that proper use of simple text can make real difference in the whole approach. Even though simple text is rarely used in terms of visualization, but our specific project we were dealing with a lot of numbers. Textually highlighting important numbers is a great option to tell story. Another way in which our design and project evolved was the kind of audience we will be targeting. Before creating our visualization product, we put a great thought to understand our audience. For example, "if our audience are technical audience who want to drill down into the analysis or non-technical audience who want to understand the high-level information?" Even though this project is a very open sourced project which will be targeting audience of both sets (technical and non-technical) it was a really important design evolution aspect that shaped our story and information accordingly and not to dwell too much into other unnecessary information. Lastly details on demand, this was by far the most important part of our whole project, providing details when needed. In our project we used this tool quite extensively just to make sure we deliver right information across to the audience. Look at the picture below to see what exactly we mean by details on demand in terms of our project.



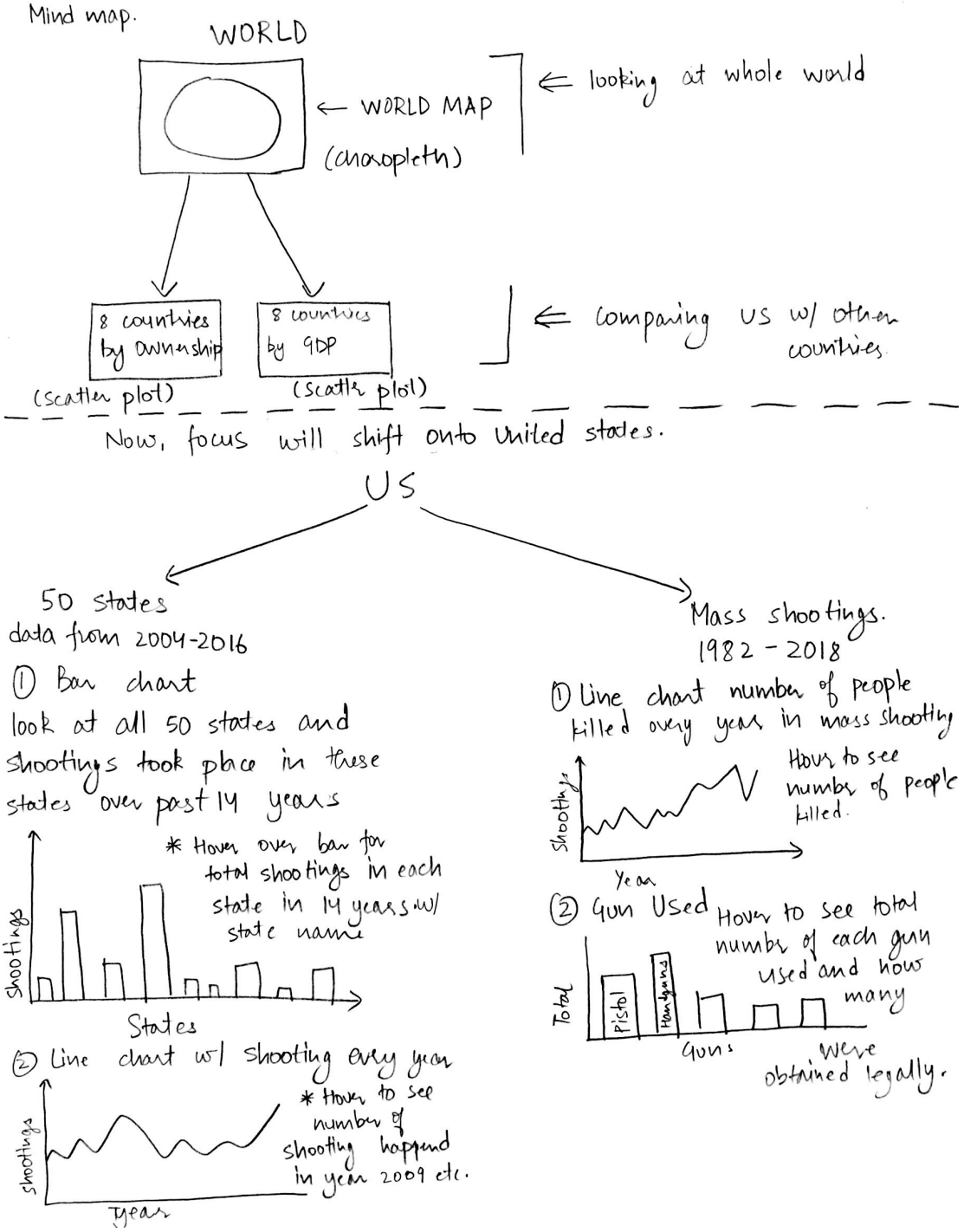
The map above is a perfect example of interaction and details on demand. Above picture is a screenshot of one of the visualization from our project. As user hover over one these circles in a given state, they will get all necessary information related with that particular circle, for example at the circle above, it gives all necessary information about Orlando night club shooting.

Overall, we did not deviated from our proposal, our project by the end became more generic and information oriented than a deep research project.

IMPLEMENTATION

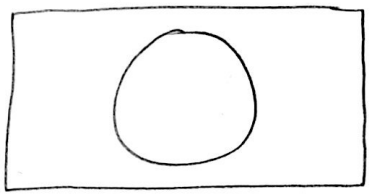
So, in order to come up with the final design we very closely followed this mind map shown below. We started with focusing on the world and then comparing United states with other countries around the world. Those comparisons were made on the basis of gun ownership and GDP for the most industrialized countries like Japan, India and Germany. After that our focus shift to US, where our project gets divided into two subparts where first part talks about general state shootings that happened in US from 2004 - 2016 and second were mass shootings that took place in US from year 1982-2018, and then finally doing our analysis on the visualizations and answering the question that we were supposed to answer through this project. Take a closer look a the mind that has been pasted in this doc to see how we implemented visualization in our project.

Mind map.



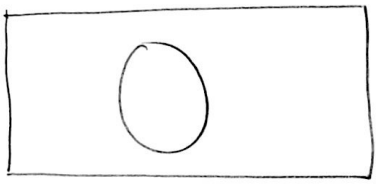
Six maps of Us about places where shootings took place.

Map 1: Showing shootings (Mass) in each state since 1982



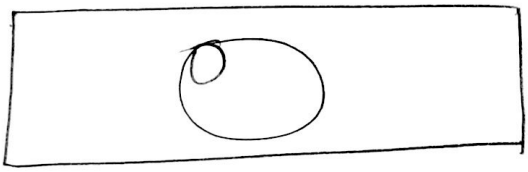
* Hover to see shootings took place in each state since 1982 + Number of people killed.

Map 2: Mass shooting in schools



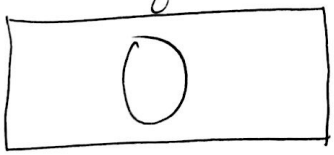
with hovers and interaction. + summary + Victims.

Map 3: Mass shooting in Workplace



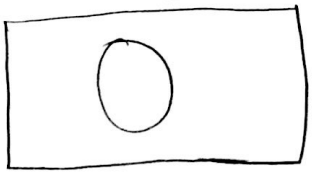
summary + victims

Map 4: Shootings at Military Bases



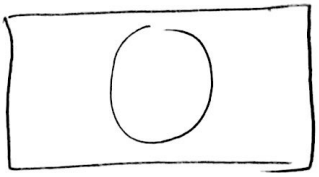
Summary + Victims

Map 5: Shootings at Religious places



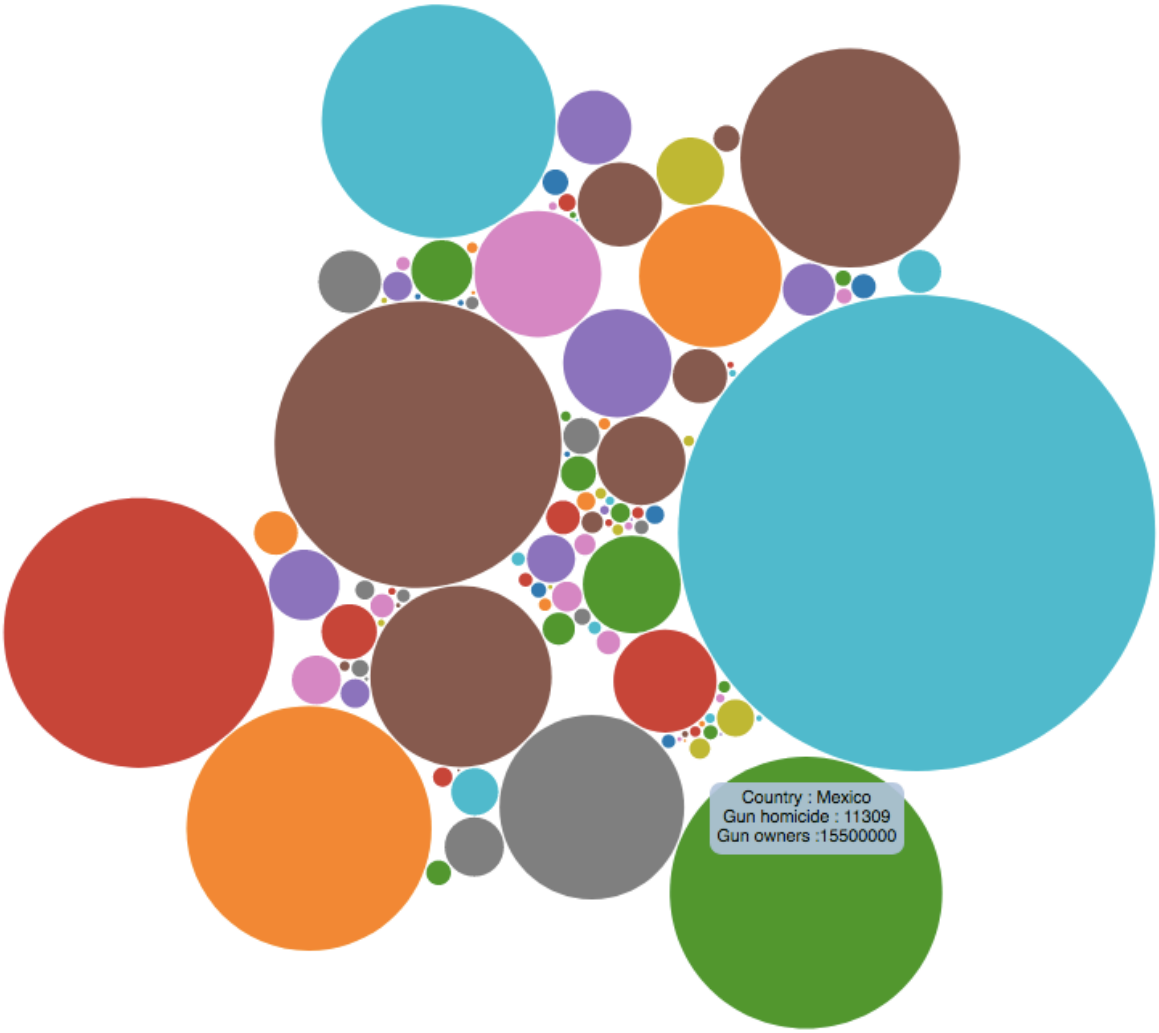
Summ. + Vict.

Map 6: Other shooting (Bars, clubs, concerts)

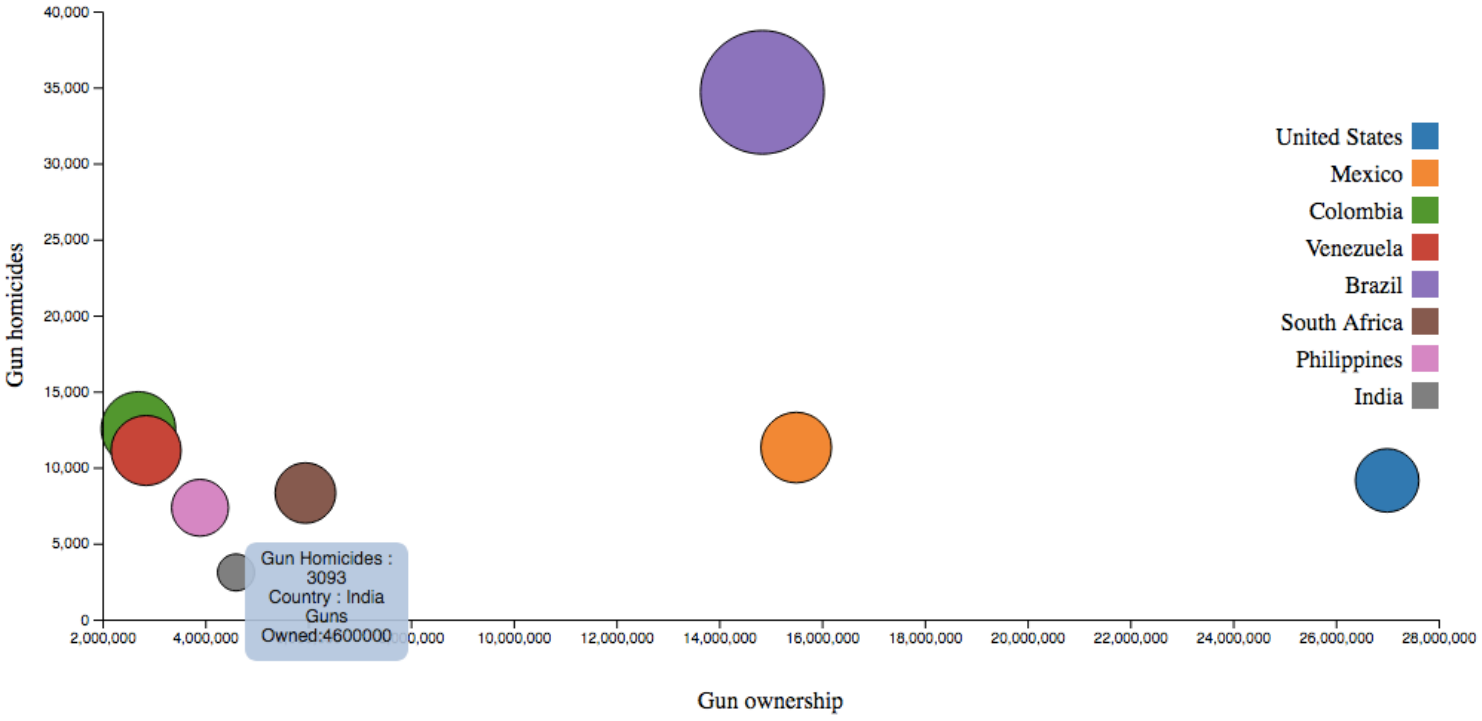


Here you will look at some of the visualizations from our project as the description related with the intent and functionality related with these visualizations.

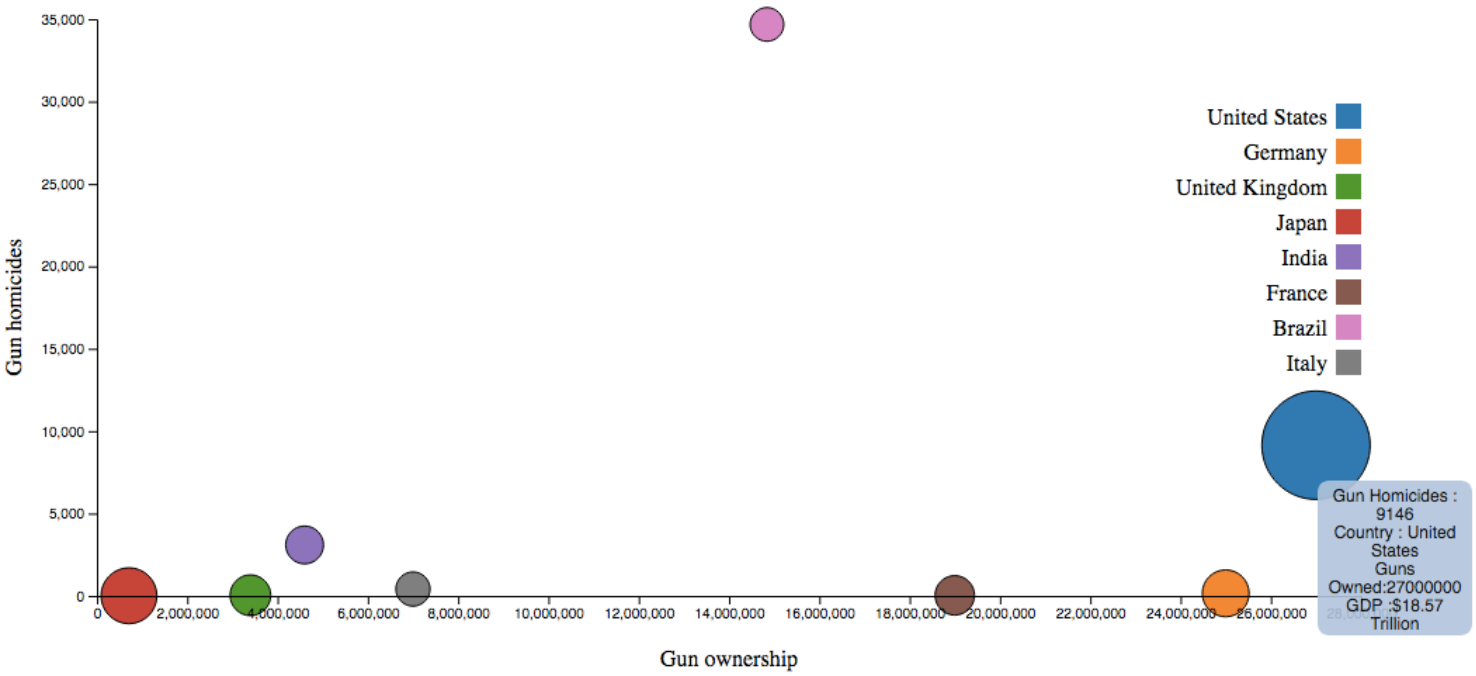
The very first picture is the picture of bubble chart with color, here as you can see that all of the countries are colored on the basis of there names. We chose this particular visualizations was because of two primary reason, first it's relatively easy to understand and second it does not shows missing data. By missing data we meant missing data values for these specific countries. For instance in this picture our cursor was on top of Mexico so it shows country name and gun homicides in this country, where as Russia is not in this bubble along side other African countries is due to unavailability of the data for these countries. Now in this picture the interaction element is the color and hover. Hovering over the country will make the country color go darker and give the user with all important information related with that specific country. This is the introductory visualization of our project.



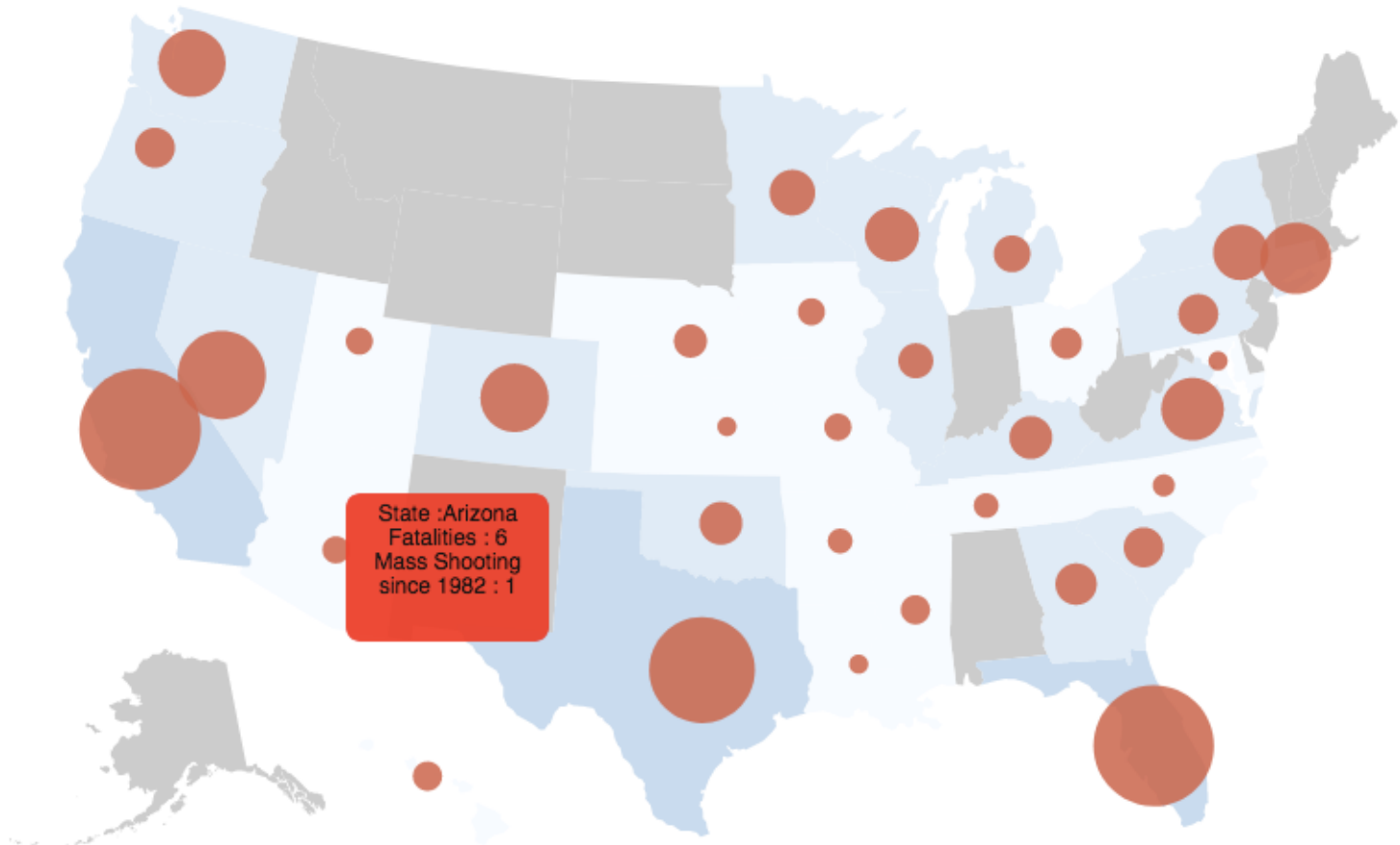
Second image we are comparing top 8 countries along side the US in terms of gun ownership and gun homicides. Here countries are colored randomly using D3's schemeCategory10 color property. X axis shows gun ownership and Y axis shows gun homicides. Size of the circle is based on the total number of gun homicide took place in these countries. Basic purpose of this chart was to see where exactly the US stands if compared with countries with high gun ownership rate and gun homicides. Interaction in this graph works as follows: Hovering over the circle will give information related with a country since each circle represents a country. For example grey circle represents India and hovering over grey circle will prompt up a dialogue box showing information related with the country



Our third graph is very similar to one above as both of them follow same principle design. In third graph we compared top 8 countries in terms of there GDP and too see how top industrialized countries are trending in terms of gun homicides and gun ownership. For this particular chart, we ended looking at GDP of each and every country separately and sizing the circle on the basis of the GDP of a country. Here interaction can be seen as we tried to hover over the blue bubble which given all necessary information related with the US. Main purpose of this chart was to rank US among other industrialized countries and to see if the US is actually doing that bad or is just media portraying it as an evil. We can clearly see where US ranks in terms of homicides by guns in terms of most developed countries across the world.



Last and fourth graph below is the choropleth map of the US and the states where mass shootings have took place since 1982. With blue color of state referring to as state where shooting took place and red circle referring total number of people killed in these mass shootings and total number of shootings that have happened in a particular state since year 1982. Clearly we can see that California has more number of mass shootings overall as compared to other states in US. Interaction in this map works in by hovering over the circle on a particular state. Simply hovering over state wont give any information but hovering over the circle will give the information regarding state as shown below about Arizona state.



The basic intent behind the functionality of the interactive visualization is to allow a reader to get appropriate information related with the data. Without interaction this project will be a big failure as it wont be able to convey all important messages and information across the board.

EVALUATION

While visualizing, cleaning and processing the data we learned a lot of different aspects related with data. From the visualizations our final opinion regarding gun violence was this that, there's surely no need to repeal 2nd Amendment as our founding fathers put in a great amount of thought while creating this amendment. However we do agree with these two aspects for sure, First there's surely a greater need of stricter gun control laws as current laws allow every person with a clear background access to guns. Some of these mass shootings took place because shooters had some sort mental illness and they obtained there weapons legally. As per our analysis number of general shootings in the US rose by 20% since year 2014 to year 2016. On other hand media in the US are portraying this country at a very bad scale in terms of gun violence. Countries like Venezuela, Brazil, Mexico are even worse. Since most of the media sources care more about political correctness than data correctness itself, we can easily convey this message through our analysis point of view that the US is not the worst country when it comes to gun violence. However, since US being the superpower and front runner in the world it does makes sense why media portrays US as the way they do right now. This analyses could go more deeper where start comparing other weapons oh homicides with gun in the US.

Overall, our visualizations work way better than what we were expecting before, however there're still some flaws. First is related with the color, to make everything run perfectly on our website we changed color property of D3 from chromatic to schemeCategory10 meaning giving random colors to everything not a fixed color (like blue or red as shown above). Apart from that most of our visualizations work just fine. For further improvement in our design we will be looking to add more user interaction and more elements in our project for instance where a user can see statistics related to each and every state over years rather than looking states all together, coming up with a generic color code so that we don't have to face color problem too much in while uploading code on the website. Lastly, further improvement in designs and story can be made by asking other people to give ratings to our visualizations and asking opinion on how to make this story and visualizations much intriguing and engaging with the reader.

WORK CITED

Below are the links to all the sources we used in order make project a success. Some of these are websites, articles, blocks with D3 code and other useful information.

(1) D3 code used to for this project, sources below :

Code for US Map

(2) D3 world Map

World Choropleth map

(3) Bubble chart for comparison

Bubble chart

(4) Small multiples and code for updating all visualizations on website

Small multiples

(5) Story/Idea for the website

Washington Post article on mass shootings

(6) Bubble chart for First graph

Bubble chart 2